DOCUMENT RESUME

ED 284 270 CS 210 717

AUTHOR	Plattor, Emma; Maguire, Thomas
TITLE	Evaluation of Selected Components of an Elementary Language Arts Program in a Small Urban School Jurisdiction. Final Report.
INSTITUTION	Grande Prairie School District No. 2357 (Alberta).
SPONS AGENCY	Alberta Dept. of Education, Edmonton.
PUB DATE	86
NOTE	281p.; For the Listening Tests: Teachers' Manuals component of this study, see CS 210 718.
PUB TYPE	Reports - Evaluative/Feasibility (142) -- Tests/Evaluation Instruments (160)
EDRS PRICE	MF01/PC12 Plus Postage.
DESCRIPTORS	Curriculum Evaluation; *Educational Assessment; Educational Research; *Educational Testing; Elementary Education; *Evaluation Methods; Foreign Countries; Holistic Evaluation; *Listening Skills; Program Content; Program Evaluation; Research Methodology; Skill Analysis; *Student Evaluation; Writing (Composition); *Writing Evaluation; Writing Research; Writing Skills
IDENTIFIERS	Canada

ABSTRACT

A two-year study designed to evaluate an elementary language arts program in the Grande Prairie School District (Edmonton, Alberta) through an assessment of student achievement was conducted from 1983 to 1985. This final report describes all facets of the study and presents the researchers' findings and conclusions. Following a preface, the report discusses the background and rationale and the research clusters and problems. The major portion of the report provides the design and findings of the study for grades one through six in the following areas: (1) the development and validation of listening achievement tests, (2) the relative value of holistic and analytic techniques for scoring written compositions, (3) teachers' evaluations of product-oriented versus process-oriented program assessments, (4) teachers' perceptions of process/product and holistic/analytic evaluation, and (5) the relationship between parents' perceptions of student achievement and product measures of actual achievement. A research summary follows the research cluster descriptions, and the final section of the report presents some conclusions and recommendations based on the results. Approximately half of the document consists of teacher marking packets for written composition evaluation in grades three and four/five; these include scoring descriptors, reporting categories, instructions, and copies of student work. (JG)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED284270

LANGUAGE ARTS PRODUCT EVALUATION

FINAL REPORT

Emma Plattor and Thomas Maguire

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

S. Wolodko

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

CS 210 717

Alberta Education

Alberta
EDUCATION

FINAL REPORT:

EVALUATION OF SELECTED COMPONENTS OF AN ELEMENTARY LANGUAGE ARTS PROGRAM IN A SMALL URBAN SCHOOL JURISDICTION

- by -

Emma Plattor

(Statistical analyses and reports
by Thomas Maguire)

Grande Prairie School District No. 2357

Grande Prairie, Alberta

Under Contract with

ALBERTA EDUCATION

1986

The contents of the Final Report: Evaluation of Selected Components of an Elementary Language Arts Program in a Small Urban School Jurisdiction, are not intended to imply, either explicitly or otherwise, Alberta Education policy regarding evaluation of student listening and written composition skills or evaluation of teaching performance.

TABLE OF CONTENTS

		PAGE
I. PREFACE		i
Steering Committee		ii
Acknowledgements		ii
II. CONTEXT OF THE STUDY		1
Background		1
Rationale		3
Purposes		7
Research Clusters/Problems		8
Importance		9
III. DESIGN AND FINDINGS OF THE STUDY		12
Research Cluster One--Listening		12
Problem One:	Listening Assessment, Grades One to Four, Discussion	13
Problem One:	Listening Assessment, Grades One to Four, Findings	35
Problem Two:	Listening Assessment, Grades Five and Six, Discussion	66
Problem Two:	Listening Assessment, Grades Five and Six, Findings	68
Problem Three:	Listening Assessment, Correlations Among Language Variables, Discussion	73
Problem Three:	Listening Assessment, Correlation Among Language Variables, Findings	73

(. . . table of contents continued)

PAGE

Research Cluster Two--Written Composition 79

Problem One: Holistic/Analytic Scoring, Grades Three to Five, Discussion 80
Problem One: Holistic/Analytic Scoring, Grades Three to Five Findings 84

Problem Two: Comparison of Holistic/Analytic Scoring, Grade Three, Discussion and Findings 94

Problem Three: Comparison of Writing Modes, Grades Four/Five, Discussion and Findings 95

Research Cluster Three--Teacher Perceptions 99

Problems One and Two: Process/Product Evaluation, Discussion 100
Problem One: Process/Product Evaluation, Findings 101
Problem Two: Process/Product Evaluation, Findings 103

Problem Three: Holistic/Analytic Comparison, Discussion 105
Problem Three: Holistic/Analytic Comparison, Findings 105

Research Cluster Four--Parent Perceptions 116

Activity One: Written Composition, Discussion and Findings 116
Activity Two: Reading, Discussion and Findings 120

IV. RESEARCH SUMMARY 123

V. CONCLUSIONS AND RECOMMENDATIONS 130

APPENDIX: TEACHER MARKING PACKAGES, GRADES THREE AND FOUR/FIVE

LIST OF CHARTS

CHART		PAGE
1	GPLT Grade Level Objectives	18
2	GPLT Stimulus Selections and Sources, Grades One through Four	19-22
3	GPLT Objectives/Items Matrix, Grades One through Four	23-26
4	GPLT Content, Grades One through Four	27-30
5	GPLT Approximate Administrative Time, Grades One through Four	31-34

LIST OF TABLES

TABLE PAGE

RESEARCH CLUSTER ONE: LISTENING

1 Grade One Item and Scale Results 38
2 Grade One Listening Test Scale Statistics 39
3 Grade One Item Factor Analysis 41
4 Norms for Grade One Listening 43
5 School/Class Scores for Grade One Listening 44
6 Grade Two Item and Scale Results 46
7 Grade Two Listening Test Scale Statistics 47
8 Grade Two Item Factor Analysis 49
9 Norms for Grade Two Listening 50
10 School/Class Scores for Grade Two Listening 51
11 Grade Three Item and Scale Results 53
12 Grade Three Listening Test Scale Statistics 54
13 Grade Three Item Factor Analysis 56
14 Norms for Grade Three Listening 57
15 School/Class Scores for Grade Three Listening 58
16 Grade Four Item and Scale Results 60
17 Grade Four Listening Test Scale Statistics 61
18 Grade Four Item Factor Analysis 63
19 Norms for Grade Four Listening 64
20 School/Class Scores for Grade Four 65
21 Correlations Among Subtests for Grades Five and Six 70

(. . . list of tables continued)

TABLE PAGE

22 Grades Five and Six Percentile Norms for Listening 71

23 School/Grade Scores for Grades Five and Six Listening 72

24 Correlations with Reading Variables, Grade Two 75

25 Correlations with External Variables, Grade Three 75

26 Correlations with External Variables, Grade Four 76

27 Correlations with External Variables, Grade Five 78

28 Correlations with External Variables, Grade Six 78

RESEARCH CLUSTER TWO: WRITTEN COMPOSITION

29 Overall District Results for Grade Three Written Composition 85

30 Correlations Among Written Composition Variables, Grade Three 85

31 Percentiles for Written Composition Variables, Grade Three 86

32 Reliability Study--Written Composition Marking, Grade Three 88

33 Overall District Results for Grade Four Written Composition 90

34 Overall District Results for Grade Five Written Composition 90

35 Correlations Among Written Composition Variables, Grade Four 91

36 Correlations Among Written Composition Variables, Grade Five 91

(. . . list of table continued)

TABLE PAGE

37 Percentiles for Written Composition Variables, Grade Four 92

38 Percentiles for Written Composition Variables, Grade Five 92

39 Reliability Study--Written Composition Marking, Grades Four and Five 93

40 Analysis of Variance, Written Composition, Grades Four and Five 97

41 Significant Interactions, Grades Four and Five 98

RESEARCH CLUSTER THREE: TEACHER PERCEPTIONS

42 Product Assessment: Benefits and Problems/Concerns 102

43 Personal/District Benefits of Product/Process Evaluations 104

44 Holistic/General Impression Marking Benefits (1983 Marking of Written Compositions) 109

45 Holistic/General Impression and Analytic Marking Benefits (1984 Marking of Written Compositions) 109

46 Comparable Benefits, Holistic/Analytic Marking 110

RESEARCH CLUSTER FOUR: PARENT PERCEPTIONS

47 Summary of Parent Marker Scoring 118

48 Summary of Teacher Marker Scoring 119

49 Summary of Parent Predictions and Actual Student Achievement on Grade Six EPSB Reading Test 122

I. PREFACE

This study, Evaluation of Selected Components of an Elementary Language Arts Program in a Small Urban School Jurisdiction, was conducted by Grande Prairie School District #2357 under contract with Alberta Education. The time span of the study was two years, from 1983 to 1985. Final data analyses and the Final Report were completed in 1986.

The study had as its major purposes: (1) the development and validation of achievement tests in the area of listening; (2) an investigation of the relative value of two techniques for the scoring of written compositions; (3) an examination of teachers' perceptions regarding the relative value of product-oriented and process-oriented program evaluation; (4) an examination of teachers' perceptions concerning the effect of their involvement in a product-oriented evaluation of the language arts program; and (5) an examination of the relationship between parents' perceptions of student achievement in selected language arts areas and actual student achievement in these areas as determined by product measures.

This Report presents a description of all facets of the study as well as the findings and conclusions of the researchers. It also includes the Teacher's Manuals and Student Test Booklets for the Grande Prairie Listening Tests (Grades One to Four) developed during the study to meet its first major purpose, and the Teacher Marking packages (Grades Three and Four/Five) developed for use in Grande Prairie School District #2357 inservice workshops on the scoring of written compositions.

The improvement of student skill in listening and written composition is a continuing goal of Grande Prairie School District #2357. This study has provided baseline data for use in determining the extent to which this goal is being reached. As well, it has culminated in instruments which district teachers can use to collect information about student strengths and weaknesses in these areas.

Such information should be of considerable value to teachers in the planning and implementing of instruction in listening and written composition. Further, involvement in the study has contributed to the professional growth of the district teachers, who can take pride in a task well done.

STEERING COMMITTEE (1983-1985)

Dr. Tom Gee, Grande Prairie Regional Office, Alberta Education

Mr. Frank Horvath, Student Evaluation Branch, Alberta Education

Mrs. Carol Onushko, Grande Prairie Regional Office, Alberta Education

Mr. Lorne Radbourne, Assistant Superintendent, Grande Prairie School District #2357

Dr. Clarence Rhodes, Planning Services Branch, Alberta Education

Dr. Harry Sherk, Planning Services Branch, Alberta Education

Mr. Keith Wagner, Deputy Superintendent, Grande Prairie School District #2357

ACKNOWLEDGEMENTS

Grateful appreciation must be expressed to the many people who contributed their time and expertise to this study. These include the teachers who served on the district's Test Development Steering Committees; the teachers who developed, piloted, revised, and finalized the test instruments; the markers who assisted in the development of the scoring criteria for the written compositions as well as the Teacher Marking Packages; and the teachers who attended inservice sessions concerning the study as it progressed.

As well, acknowledgement must be given to the Grande Prairie Regional Office consultants who helped with the test development and with the writing of the Final Report, university consultants including Drs. Tom Maguire and Pat Payne of the University of Alberta, Dr. Peter Evans of the University of Victoria, and Dr. Emma Plattor of the University of Calgary; secretarial staff of both the Grande Prairie School District #2357 and the Department of Teacher Education and Supervision of the University of Calgary; and the district school administrators who facilitated the implementation of the study.

A very special thank-you must go to Derek Taylor, Superintendent of Schools, Grande Prairie School District #2357, who believed in the study and supported it fully. Appreciation must also be expressed to Keith Wagner, who initiated the study during his term as Deputy Superintendent, and Lorne Radbourne, Assistant Superintendent, who continued its development when Keith Wagner left Grande Prairie to assume duties with Alberta Education. Without these Central Office administrators, the study might not have come to fruition.

Finally, special thanks must also be given to Drs. Clarence Rhodes and Harry Sherk of Alberta Education's Planning Services Branch, who facilitated the approval of the study and guided its development.

A project of this breadth and scope is only as strong as the people who have faith in it and are willing to stay with it all the way. So many people did just that, and the result has been a vibrant and viable study which will affect students positively for years to come.

II. CONTEXT OF THE STUDY

BACKGROUND

Grande Prairie School District #2357 is a public school district whose boundaries are coterminus with those of the city of Grande Prairie, Alberta. The city's population is approximately 25,000 with about 4,000 students in its eight schools. One junior high school houses Grades Seven to Nine, and one senior high school houses Grades Nine to Twelve. Two of the six elementary schools contain junior high school grade levels. In the 1984-85 school year, there were approximately 1700 students in the district's elementary schools, and about 74 teachers at these levels, excluding special personnel of any type. There are no subject area consultants or supervisors in the Central Office component. The Assistant Superintendent for Educational Programs serves in these capacities. The district is classified as a small urban school jurisdiction.

Beginning in the mid 1970s, the district developed a plan whereby program areas (e.g.: elementary language arts, junior high math, senior high social studies) were selected for evaluation on a rotating basis, with the intent of evaluating every program every five years. In any given year thereafter, the district might have been involved in four or five such evaluations at the various divisional levels. These program evaluations were designated "process evaluations"--that is, they examined what might best be termed inputs and processes. By inputs were meant such aspects as teaching resources, materials, and the like. Processes meant teaching strategies, activities, and the like, and were investigated through direct classroom observations.

A process evaluation generally involved an on-site visit of about a week's duration by a team which usually included a member of an Alberta Education Regional Office, a member of a university faculty, a school administrator from another system, and at least one teacher from another system, all of whom were assumed to have some special knowledge in the subject area under investigation. The team did not examine student achievement in any depth, but rather examined resources and observed classroom teaching. The result of the investigation was a report to the district which included a program description and recommendations for growth and change. This report would then become an agenda for change.

As defined by Grande Prairie School District #2357 for the study described in this Report, "product evaluation" involves the assessing of student outcomes--evaluating student achievement. It is concerned with the products that the students provide. Thus it is based on the monitoring of achievement, and focusses on what students produce rather than on instructional processes used by teachers. Since definitions of product and process vary, it is important that this distinction be recognized.

Product assessment had of course been in effect in the district--students were tested with standardized instruments obtained from commercial sources, other school districts, Alberta Education, and the like. But this type of assessment had not previously been used to evaluate programs. Thus one of the thrusts of this study was to examine the feasibility for this type of program assessment.

RATIONALE

The Proposal (1983, pp. 1-3) for this study contained the following Rationale Statement.

> At the present time there appears to be a great deal of public concern with the product outcomes of schools. Alberta Education has reacted to that concern by instituting compulsory measures of product on periodic bases in selected areas, one of which is language arts. At the local level, further product measures, including attempts to correlate perceptions of achievement levels with measured achievement, might help to restore public confidence.
>
> At the school system level, there is a need for achievement monitoring that is more regular and grade specific than occurs at provincial levels. For example, at provincial levels, language arts achievement testing occurs only in Grades Three and Six and only every fourth year. Teachers and school systems find this of limited value, and would prefer assessments that are more immediate and more closely related to determining achievement levels in specific objectives as defined by the system within the broad goal parameters set out by the Alberta Education curriculum.
>
> Provincial efforts in measuring achievement in language arts appear to be emphasizing reading and writing. However, listening and speaking are important components of a language arts program. Inasmuch as product measures may have a determining effect on curriculum, to omit these components from evaluation efforts may have the effect of encouraging teachers to continue to downplay their importance. In its process-oriented school evaluations, Alberta Education has (in the past) persistently commented on the lack of

attention to the listening dimension, both in teaching emphasis and evaluation emphasis. For a small school district to undertake efforts in the assessment of listening and speaking, both of which are uncharted areas, at one time would be impractical; hence this study proposes to focus on only the listening dimension of oracy. Since another district, Edmonton Public, has already done some work at the Grades Five and Six levels, this proposed study will focus on the development of listening assessment materials in Grades One through Four. The Edmonton Public Listening Tests at the Grades Five and Six levels will be used concurrently with this study, and further research into domains of listening will be explored through factor analyses.

In the literacy dimension of language arts, the assessment of reading skills has received extensive study. Commercial firms and large school districts such as Edmonton Public have produced some excellent tests. Once again it is intended that the Edmonton Public Reading Tests would be used concurrently with this study but findings will not be requested as an integral part of the study. However, the data base provided will permit studies examining the correlation of reading achievement and achievement on listening tests and tests of ability in written composition.

The evaluation of the written composition aspect of the literacy dimension has yet to receive the attention that reading has received, especially at the elementary level. In the United States, the NAEP (National Assessment of Educational Progress) Commission has laid some groundwork in the application of holistic and primary trait scoring techniques to the written composition of students aged 9, 13, and 17.

The following statement indicates that NAEP does not consider their work to be the final story:

> Papers written in response to tasks requesting a discourse on feelings and ideas were written for the purpose of being expressive. They agreed that procedures for rating the success of explanatory and persuasive writing could be defined since the desired effect is usually clear. On the other hand, the definition of success in the third area is more subjective in nature, since the intended or appropriate audience for writing alone for expressive purposes is less clear. (Mullis, 1980)

Once again, the Edmonton Public School Board has produced a useful background document entitled Evaluation: Responding to Children's Writing. However, it is not based on research findings. Nyberg and Nyberg (1982) have conducted Alberta based research exploring the validity and reliability of holistic scoring techniques at the Grade Twelve level, and have published a useful research summary which includes essays representative of various stanine gradings. Little is yet known about the reliability of such techniques at the elementary level. Similar research and the publication of a document similar to Nyberg and Nyberg's would be useful.

In regard to the evaluation of written composition, an important issue that has received little investigation is the effect of mode or type of writing assigned on the rating of quality. It has been determined by Crowhurst (1977) and by Wagner (1983) that mode of writing can have an effect on such components of writing quality as syntactic complexity.

A study by Prater and Padia (1983) indicates that expressive writing tasks generated higher quality essays than explanatory or persuasive writing tasks at the fourth and sixth grade levels.

Further study of this nature is warranted, such as that of Alberta Education in assessing writing competency in Grades Three and Six.

Whenever any type of program evaluation is carried out, there is potential for the assessment of that evaluation on those delivering the program, in this case the teachers. Over the past few years teachers in the Grande Prairie School District have been the object of several process-oriented evaluations. The evaluation activities associated with this study are product-oriented. Thus it would be of interest to study the perceptions of teachers regarding the relative values of these types of evaluations. In as much as Alberta Education has been and continues to be engaged in process-oriented evaluations, and has recently placed increased emphasis on product evaluation, such findings should be of interest.

As indicated at the outset, much of the concern about product in language arts arises because of public perception. Parents and other interested citizens seem to feel that student ability in language arts is not what it should be. At times, the public is very specific in its concerns; for example, there appears to be a perception that phonetic word attack skills are not adequately taught. However, there has been little attempt at correlating public perception with the actual demonstrated achievement of students.

This study proposes to study these correlations, albeit in a somewhat limited way. Once again, the findings should be of general interest. Perhaps the public concern relates more to their reaction to media articles than to their reaction to demonstrated student achievement. If so, increased public relations efforts should be mounted.

It is proposed that all of the issues addressed above can be explored in one rather comprehensive study.

PURPOSES

The study had five major purposes. In developing the design of the study, each purpose was represented within a research cluster with its set of specific questions. These major purposes were:

1. the development and validation of achievement tests in the area of listening;
2. an examination of the relative value of two techniques for the scoring of written compositions;
3. an examination of teachers' perceptions regarding the relative value of product-oriented and process-oriented program evaluation;
4. an examination of teachers' perceptions concerning the effect of their involvement in a product-oriented evaluation of the language arts program;
5. an examination of the relationship between parents' perceptions of student achievement in selected language arts areas and actual student achievement in these areas.

As well, the study had several secondary purposes related to those listed above. These were:

1. an examination of the correlations between listening skills and other selected language arts skills;
2. an investigation of the empirical validity of the conceptual domains upon which the Edmonton Public Listening Tests (Tests of Listening Comprehension) were based;
3. an investigation of achievement in written composition based on different writing modes;
4. an investigation of the effect on teaching emphases and behavior of involvement in a product-oriented evaluation.

RESEARCH CLUSTERS/PROBLEMS

In order to meet the purposes of the study, four research clusters were specified, each with its own set of problems.

Research Cluster One--Listening

1. Can student achievement in the listening dimension of the Alberta Elementary Language Arts Curriculum be assessed validly and reliably at the Grades One to Four levels?
2. Do the conceptual domains of listening skills which underlie and form the basis of the construction of the Edmonton Public Listening Tests have empirical validity?
3. What other language arts skills correlate most highly with listening skills?

Research Cluster Two--Written Composition

1. Can the analytic scoring techniques that have been applied at secondary school levels be used reliably for the rating of written compositions at Grades Three, Four, and Five?
2. How does analytic scoring compare to holistic scoring in Grade Three in each of the following dimensions: (a) inter-rater reliability; (b) cost effectiveness; (c) teacher satisfaction (both in terms of those teachers involved in scoring and those who receive the information from scoring)?
3. Are there significant differences in achievement in written composition in Grades Four and Five dependent on the mode of writing required?

Research Cluster Three--Teacher Perceptions

1. What is the effect, as reported by teachers, of involvement in a product evaluation, on teaching emphases and behavior?
2. What is the perception of teachers regarding the relative values of a product-oriented evaluation and a process-oriented evaluation?
3. What is the perception of teachers regarding the relative values of holistic and analytic scoring of written composition?

Research Cluster Four--Parent Perceptions

1. What is the relationship between parental perceptions of student achievement in selected areas of language arts, and the actual achievement as determined by product measures?

IMPORTANCE

A study of this nature can result in considerable benefit not only to the teachers and students in the district in which it is conducted, but also to teachers and students in other districts and to the public at large. A number of these benefits were specified in the original Proposal.

Concerning the area of listening, potential benefits include: (1) provision of a pool of test items for use in the development by other jurisdictions of listening tests at the elementary level; (2) development of test instruments based on the Alberta language arts curriculum; (3) collection of information about the nature of the listening task as well as the domains and skills which comprise this task; (4) provision of information about student needs in the various listening skill areas; and (5) guidance for the planning and implementing of the instructional program in listening.

The latter benefit is of particular importance in an area where teachers have had very little instruction at either the preservice or inservice level. Two quotes from the literature underscore this point. Swanson (1986) points out that teachers need to model good listening skills but that few do. As he puts it:

> Most teachers are not trained in listening. They tend to be operating from the "flat earth" perspective of the 3Rs, believing that listening develops naturally, without instruction. Unfortunately, that natural development assumption is true for very few. (p. 10)

In 1977, Shuman published a book, Questions English Teachers Ask, which provided some of the answers of over 3,000 teachers to a letter requesting that they send to him questions which perplexed them. Shuman tried to balance the sections of his book in a proportion roughly resembling the categories in which the teachers asked their 946 questions. He wrote:

> For example, I would have liked to have major sections such as drama, speaking, listening, and articulation among secondary schools, colleges, and universities. However, because only 11 questions had to do with speaking and listening, I had to conclude that these topics were not foremost in teachers' minds. (Preface, unpaged)

It is also possible to conclude that teachers do not know which questions to ask about speaking and listening--that if they did they would ask them! Whatever the conclusions, the result is the same--teachers need help with program planning in listening.

Concerning the area of written composition, potential benefits include: (1) provision of information concerning the reliability of analytic and holistic scoring procedures at elementary grade levels; (2) collection of information concerning the relative costs and usefulness to teachers of these scoring techniques; (3) development of a set of descriptors for analytic marking scales for use at the elementary grade levels, as well as a handbook containing a description of the process, the descriptors, and representative compositions; and (4) provision of information about the effect of writing mode (in this case narrative, descriptive, and persuasive) on achievement in written composition.

Concerning the area of product evaluation versus process evaluation, potential benefits include the determining of the relative values of these two techniques and the concomittant provision of this information to other jurisdictions who have similar concerns as to the appropriateness and utility of these types of program assessment. If educators are to be able to design appropriate evaluations, especially in the area of program and curriculum, they must be able to clarify the purposes of different evaluation strategies as well as specify their goals and implementational components. Consideration of the relative merits of process and product evaluations should go some distance toward developing skill in doing these things.

Concerning the area of parental perceptions of student achievement, potential benefits include: (1) provision of direction to the school district regarding parental expectations; (2) collection of information concerning the need for improved or alternative public relations techniques; and (3) provision of information concerning the relationship of teacher perceptions and parent perceptions about student writing skill. The closer that parents and teachers can come in their perceptions about student achievement the more likely it is that conflicts can be avoided.

Perhaps the most important benefit, or at least a benefit as important as those listed above, is the potential for teacher staff development and change through involvement in a project of this nature. As teachers take ownership in the project, and as they find answers to questions that they raise as their knowledge grows, the likelihood of positive change should increase. As teachers and university consultants and facilitators interact on a collaborative basis, the doors to change should open wider. Smulyan (1984) notes that collaborative inquiry contributes to both knowledge in the field and improved practice since it can provide teachers with the opportunity to gain knowledge and skill in research methods and applications, become more aware of options and possibilities for change, and become more critical and reflective of their own practice. She also suggests that, if teachers work together on a common problem, communicating and clarifying ideas and concerns, they will be more likely to change their attitudes, beliefs, and behaviors if their own research findings indicate that such change is necessary.

III. DESIGN AND FINDINGS OF THE STUDY

In the Fall of 1983, an application was made and formally approved for funding by Alberta Education for a study titled: Evaluation of the Elementary Language Arts Program in Small School Jurisdictions Through Product Assessment. A Steering Committee was formally appointed consisting of members of Alberta Education and Keith Wagner, then Deputy Superintendent in Grand Prairie School District #2357. This Steering Committee held its first meeting in late November, 1983.

In order to meet the purposes of the study, four research clusters were developed, each consisting of (1) a set of problems in the form of questions to be answered through the study; (2) an action plan and timeline; (3) a description of possible benefits; and (4) a budget. These clusters were labelled Listening, Written Composition, Teacher Perceptions, and Parent Perceptions.

Prior to the formalization of the study, considerable preliminary work had been done in the areas of listening and written composition. A brief description of this work will be included with the discussion of each research cluster.

RESEARCH CLUSTER ONE: LISTENING

The following problems concerning the area of listening were presented for investigation in the study.

1. Can student achievement in the listening dimension of the Alberta Elementary Language Arts Curriculum be assessed validly and reliably at the Grades One to Four levels?

2. Do the conceptual domains of listening skills which underlie and form the basis of the construction of the Edmonton Public Listening Tests have empirical validity?

3. What other language arts skills correlate most highly with listening skills.

Two of the purposes for the study underlie the development of these questions. These were (1) the development and validation of achievement tests in the area of listening; and (2) an examination of teachers' perceptions regarding the relative value of product-oriented and process-oriented program evaluation.

In order to satisfy the second purpose, it was necessary to develop and validate achievement tests in listening, at least at the Grades One through Four levels, where no appropriate instruments existed. At the Grades Five and Six levels, an instrument Canadianized by the Edmonton Public School Board was available, but there was some concern as to its structure and thereby to its utility. As well, there was considerable interest on the part of teachers concerning the relationship of listening skills with other language arts skills (e.g.: reading, writing).

Problem One: Listening Assessment, Grades One to Four, Discussion

The following Action Plan and Timeline for this problem were included in the Proposal.

(1) June 83	Specify listening objectives at each of Grades One to Four.
(2) April 83	Design test items relating to the specified objectives; design test blueprint and specify conceptual domains.
(3) June 84	Administer test items.
(4) July 84	Conduct statistical item analyses.
(5) November 84	Design pilot listening tests for Grades One to Four.

(6) February 85	Administer pilot tests; conduct statistical analysis of reliability and factor validity of conceptual domains.
(7) April 85	Revise tests as indicated.
(8) June 85	Administer revised tests.
(9) July/August 85	Further analyses of revised tests.

In general, the Action Plan and Timeline were reasonably well followed. By June, 1983, not only had the objectives been specified but the first version of a set of listening tests had been developed and administered, one at the Grade One level and the other at the Grade Two/Three level. Test domains and objectives were generally based on those developed for the Alberta Listening Tests, which in turn had been developed for the Minister's Advisory Committee on Student Achievement (MACOSA) and published in 1979. However, only a Grade Three and a Grade Six version of the MACOSA tests were available, so objectives had to be specified for the Grade One and Grade Two/Three versions of the Grande Prairie tests. These were administered to all Grade One, Two, and Three classes in June, 1983. At the same time, the Edmonton Public Reading Tests were administered to Grades One to Six and the Edmonton Public Spelling Tests were administered to Grades Two to Six.

The conceptual domains for this first version of the listening tests were specified as Literal Comprehension, Inferential Comprehension, and Intonation. Intonation was listed in the Teacher's Manual as a subset of Inferential Comprehension. In later versions of these tests, it was included as one area of Inferential/Critical Comprehension.

The subtests were developed in terms of stimulus selections and language types. This was consistent with the development of the Alberta Listening Tests (MACOSA). The Grade One listening test involved subtests labeled "Real Versus Make Believe," "Sentences With Similar Meanings," and "Intonation." The Grade Two/Three Test consisted of four subtests, one concerned with student talk (a recipe), one with teacher talk (a field trip), one with the reading of a story, and one with a conversation in the form of a monologue and then a dialogue. Thus the language types were spontaneous spoken language and written language read aloud.

In a review of the literature on listening assessment, Plattor (1977) had noted the following recommendations: (1) Listening assessments should be based on specified objectives for listening instruction; (2) Listening comprehension should be tested under varying conditions of natural speech situations; and (3) In order to measure communicative competence in oral language, tests should account for context, audience, and purpose. Both the Alberta and Grande Prairie tests took these recommendations into account. Further, the Alberta Listening Tests (MACOSA) were concerned with including both spoken language and written language read aloud, and accomplished this through varying the subtest content for this purpose. Again, the Grande Prairie tests followed this precedent.

It may be useful at this point to explain the emphasis on these types of language and listening formats. The concern with "natural spoken language" is basic to assessment of listening. It would appear that there are two types of spoken language--spontaneous and prepared. Spontaneous spoken language is generated as the speaker talks and is organized as the speaker's thoughts take shape. On the other hand, prepared spoken language is organized prior to its use in a communication situation (as in giving a report from notes) and is generally in accord with a predetermined purpose and audience. Both spontaneous and prepared spoken language may reflect casual, informal, and formal levels of language use. (Wilkinson et. al., 1974)

However, both these forms of spoken language are different from written language in several ways. In written language, sentences are carefully structured and linked. Written language is tightly organized and contains more objective details than does spoken language. Spoken language is often loosely organized and may be characterized by incomplete or ungrammatical utterances and repetitive false starts. Written language contains punctuation to indicate such vocal factors as pitch, intonation, and stress. Thus an oral reader must follow the writer's vocal clues. Vocabulary and writing conventions must also be used to provide information given by the speaker through vocal and nonverbal factors. (Plattor et. al., 1979) Thus attention must be given in listening assessment to both spontaneous and written language read aloud. Spontaneous prepared language was included only in the Grade Four test, developed later.

Statistical analyses of the first versions of the Grande Prairie Listening Tests suggested certain improvements and extensions of the Grade One test. The Grade Two/Three test was found to be too easy for the Grade Threes. It was decided to keep the test for use in Grade Two, to redevelop the Grade One test, and to build new tests for the Grades Three and Four levels.

During the Fall of 1983 and the Winter of 1984 test development continued in this regard. As well, consultants from both the University of Calgary and the University of Alberta met with teachers concerning the test development as well as for listening inservice workshops. However, the second version of the tests was not completed by June, 1984, as planned. The task was larger than had been anticipated, and teachers-developers did not wish to sacrifice quality to time constraints.

The second version of the tests, now titled Grande Prairie Listening Tests (GPLT), was completed and administered in October, 1984. In order to maintain test administration consistency, the Grade Two test was given to Grade Three, the Grade Three test to Grade Four, and the Grade Four test to Grade Five. Again, statistical analyses were conducted, results studied, and revisions made.

Version C, the final version of the GPLT, was administered to all Grades One to Four students in June, 1985. This involved 289 students at Grade One; 291 at Grade Two; 270 at Grade Three; and 265 at Grade Four.

At each level of the GPLT, the subtests reflect spontaneous language as well as written language read aloud. Each subtest consists of a stimulus selection and a set of questions about the selection. Three alternative answer choices are provided for each question. In general the contexts of these stimulus selections are classroom based, although at the Grade Four level some are more public (weather and news reports). The selections also represent both teacher and student spoken language.

Each of Versions A, B, and C of the tests involved the development of Teacher's Manuals, Student Test Booklets, and audio tape recordings of the entire content of the tests at each level. Directions for administering the tests were provided in each Teacher's Manual, along with the entire tapescript, again at each level. All tests were administered in regular classrooms by the classroom teachers. Answers were provided in the Teacher's Manuals.

Chart 1 presents the objectives by grade level for the final version (Version C) of the GPLT. Chart 2 lists the subtest stimulus selections and their sources for each grade level. Chart 3 provides an Objectives/Items Matrix for each grade level. Chart 4 provides the test content for each grade level, including the stimulus format, language type, number of items, and approximate time of each subtest. Chart 5 provides the approximate total administrative time at each grade level.

CHART 1

GPLT Grade Level Objectives

Objectives	Grade Level 1	2	3	4
Literal Comprehension				
1. Identifying main idea	X	X	X	X
2. Identifying supporting details	X	X	X	X
3. Matching/comparing details	X	X	X	X
4. Identifying sequence of details	X	X	X	X
5. Identifying word meanings	X	X	X	
6. Following directions				X
Inferential/Critical Comprehension				
1. Making judgements/drawing conclusions	X	X	X	X
2. Inferring feelings	X	X	X	X
3. Predicting details/outcomes	X		X	
4. Noting similarities and differences				X
5. Relating intonation to feelings	X		X	
6. Identifying cause and effect				X
7. Identifying point of view				X

CHART 2

GPLT Stimulus Selections and Sources, Grade 1

Subtest	Content	Source
1. Student talk: Recipe	"Wonder Soup"	G.P.S.D. #2357
2. Teacher talk	"Circus"	G.P.S.D. #2357
3. Story	Martin Levin "An Elephant Tale"	SRA Individualized Reading Skills Program, Strawberry Emergency Science Research Associates (Canada) Limited 1974, pp. 102-105
4. Conversations	A. Cathy's mother	Original
	B. "Freddy Frog and Meadow Mouse"	"A House for Googie" John A. McInnes et al. Heads and Tails, Thomas Nelson and Sons (Canada) Ltd., 1977, pp. 12-13

CHART 2--continued

GPLT Stimulus Selections and Sources, Grade 2

Subtest	Content	Source
1. Song	"The Back of the Crocodile"	Sharon, Lois, and Bram Elephant Jam, McGraw-Hill Ryerson, 1980 p. 26
2. Poem	"The Crocodile's Toothache"	Shel Silverstein, Where the Sidewalk Ends. Harper and Row, 1974, p. 66
3. Student Interview	"Saturday Morning Pet Parade--Second Interview"	John McInnes et. al. Saturday Magic, Toronto: Nelson Canada, 1977, pp. 45-48
4. Picture Stories	Original	G.P.S.D.# 2357

CHART 2--continued

GPLT Stimulus Selections and Sources, Grade 3

Subtest	Content	Source
1. Student talk: Recipe	"Wonder Soup"	Original Story, G.P.S.D. #2357
2. Story	"Story 42"	Adapted from a story by Stephen Southwold in Developing Comprehension in Reading, Level-5, J.M. Dent and Sons (Ltd), 1963, pp. 191-197.
3. Poem	William Wise "After the Party"	Bill Martin Jr. and Peggy Brogan. Sounds After Dark. Holt, Rinehart, and Winston, 1972 p. 288.

CHART 2--continued

GPLT Stimulus Selections and Sources, Grade 4

Subtest	Content	Source
1. Short Story	"The Monkey and the Lion" by Denney Dey and Gary Grimm	Storytelling 1979; p. 17-18
2. Poetry	"The Cave Boy" by Laura E. Richards	Starting Points in Language, Book 2, Teachers' Guide, Ginn, 1974
3. Following Directions	Original	G.P.S.D. #2357
4. News Report	"Space Shuttle Landing" Gord Sharp	Grande Prairie Radio Station, April 11, 1984 8:00 am News
5. Weather Report	Weather Report	Grande Prairie Radio Station, April 11, 1984
6. Class Discussion	"Hockey schools help improve the game of hockey"	Teaching Strategies Sourcebook 2. Gage, 1973

CHART 3

GPLT Objectives/Items Matrix, Grade 1

Objectives	Subtests/Items			
	Student Talk: Recipe	Teacher Talk: Circus	Story	Conversations
Literal Comprehension				
1. Identifying main ideas	1	2		
2. Identifying supporting details		3,6		
3. Matching verbal details	4	5		
4. Identifying sequence of details			2	
5. Identifying word meanings			1	
Inferential/Critical Comprehension				
1. Making judgements	2	1	3,5	
2. Drawing conclusions	3		6	
3. Inferring feelings		4	4	
4. Predicting details in sequence				4,6
5. Relating intonation to feelings				1,2,3,5

CHART 3--continued

GPLT Objectives/Items Matrix, Grade 2

Objectives	Subtests/Items			
	Song	Poem	Interview	Picture Stories
Literal Comprehension				
1. Identifying main ideas		7		
2. Identifying supporting details	2	8, 11	13, 14	
3. Matching verbal and pictorial details				17, 18, 19 20, 21
4. Identifying sequence of details	6	9		
5. Identifying word meanings	3		15	
Inferential/Critical Comprehension				
1. Making judgements		10	16	
2. Inferring feelings	1, 4, 5	12		

CHART 3--continued

GPLT Objectives/Items Matrix, Grade 3

Objectives	Subtests/Items		
	Recipe	Story	Poem
Literal Comprehension			
1. Identifying main ideas	1	1	2
2. Identifying supporting details			3
3. Identifying sequence of details		2,3	
4. Comparing details	4	9	
5. Identifying word meanings		6	
Inferential/Critical Comprehension			
1. Predicting outcomes	3		5
2. Making judgements	2	5,10	
3. Drawing conclusions		4,7	1,6
4. Relating intonation to mood and meaning		11,12	
5. Inferring feelings		8	4

CHART 3--continued

Objectives/Items Matrix, Grade 4

Objectives	Subtests/Items					
	Short Story	Poem	Following Directions	News Report	Weather Forecast	Classroom Discussion
Literal Comprehension						
1. Identifying main idea		7		18		25
2. Identifying supporting details	1	10		12,14, 15,16, 17	19,20, 21,22,	28
3. Identifying sequence of details	4,5					
4. Comparing details		9			23	
5. Following directions			11			
Inferential/Critical Comprehension						
1. Noting similarities and differences						
2. Identifying cause and effect				13		30
3. Identifying point of view						27
4. Drawing conclusions, generalizing	2,3				24	29
5. Inferring feelings	6	8				26

CHART 4

GPLT Content, Grade 1

Subtest	Stimulus Format	Language Type	Items	Total Number of Items	Percent of Items	Approximate Time (min.)
	Instructions					4
1.	Student Talk: Recipe	Spontaneous	1 - 4	4	19	4
2.	Teacher Talk: Circus	Spontaneous	1 - 6	6	27	6
3.	Story	Written	1 - 6	6	27	8
4.	Conversation	Spontaneous	1 - 6	6	27	5
				22	100	27

CHART 4--continued

GPLT Content, Grade 2

Subtest	Stimulus Format	Language Type	Items	Total Number of Items	Percent of Items	Approximate Time (min.)
	Intro.					2
1.	Song	Written	1 - 6	6	29	3
2.	Poem	Written	7 - 12	6	29	4
3.	Student Interview	Spontaneous	13 - 16	4	19	3
4.	Picture Stories	Written	17 - 21	5	23	3
				21	100	15

CHART 4--continued

GPLT Content, Grade 3

Subtest	Stimulus Format	Language Type	Items	Total Number of Items	Percent of Items	Approximate Time (min.)
	Instructions					2
1.	Recipe	Spontaneous	1 - 4	4	18	5
2.	Story	Written	1 - 12	12	55	13
3.	Poem	Written	1 - 6	6	27	5
				22	100	25

CHART 4--continued

GPLT Content, Grade 4

Subtest	Stimulus Format	Language Type	Items	Total Number of Items	Percent of Items	Approximate Time (min.)
	Instructions					1
1.	Short Story	Written	1 - 6	6	20	8
2.	Poem	Written	7 - 10	4	13	4
3.	Following Directions	Written	11	1	04	2
4.	News Report	Planned Spontaneous	12 - 18	7	23	4
5.	Weather Forecast	Planned Spontaneous	19 - 24	6	20	2
6.	Classroom Discussion	Spontaneous	25 - 30	6	20	7
				30	100	28

CHART 5

GPLT Approximate Administration Time, Grade 1 *

Grade	Preparation and Instructions	Tape Content (Minutes)	Total Time (Minutes)
1	20	27	47

---

* A break after Subtest 2 may be taken at the discretion of the test administrator.

CHART 5--continued

GPLT Approximate Administration Time, Grade 2 *

Grade	Preparation and Instructions	Tape Content (Minutes)	Total Time (Minutes)
2	20	15	35

---

* A break after Subtest 2 may be taken at the discretion of the test administrator.

CHART 5--continued

GPLT Approximate Administration Time, Grade 3 *

Grade	Preparation and Instructions	Tape Content (Minutes)	Total Time (Minutes)
3	15	25	40 (plus break if desired)

---

* A break after Subtest 2 may be taken at the discretion of the test administrator.

CHART 5--continued

GPLT Approximate Administration Time, Grade 4 *

Grade	Preparation and Instructions	Tape Content (Minutes)	Total Time (Minutes)
4	12	28	40 (plus break if desired)

---

* A break after Subtest 2 may be taken at the discretion of the test administrator.

Problem One: Listening Assessment, Grades One to Four, Findings

Statistical analyses for the final version (Version C) of the Grande Prairie Listening Tests were conducted by Dr. Tom Maguire and the results presented by him in a report titled: The Evaluation of Instruments Designed to Assess the Listening Dimension of the Alberta Language Arts Curriculum, Grades One to Four. This report is presented below in its entirety.

The report's Summary provides the answer to the first problem in Research Cluster One--Listening: Can student achievement in the listening dimension of the Alberta Language Arts Curriculum be assessed validly and reliably at the Grades One to Four levels?

The answer to this question is a qualified yes. The tests developed for Grades One and Two are both good tests and could be used for identifying students for further diagnostic work. In addition, they could be used for program monitoring. Subscale or passage scores do not have validity by themselves, but at the program level it would be appropriate to use item scores.

The Grade Three test requires further investigation to see if the practice effect found (adjusted test-retest reliability) was an anomaly, or if it is a general phenomenon. If further investigation supports the former hypothesis, then the test could be used for the purposes described for Grades One and Two.

The Grade Four test is the least successful of the instruments. Further research is needed to see if this is a property of the test or of general student development. In particular, given the compelling face validity, it would be sensible to pursue investigations into multiple factors such as following directions for verbal products, following directions for spatial products, attending to details, interpreting, and extrapolating.

Instrument Evaluation Report (Maguire, 1986)

Introduction

In 1983, the Grande Prairie School District undertook the development of tests to measure the listening objectives of the Alberta Elementary Language Arts Curriculum in Grades One to Four. Two cycles of development and revision were undertaken to produce instruments suitable for students in these grades. An evaluation of the instruments in their final form is provided here.

The following sections of this report have been organized by grade. Within each grade, information on the internal aspects of the tests is provided (item quality, item structure, and test reliability), followed by evidence relating to the external validity of the test (school and class discriminations, correlations over time, and correlations with outside variables).

Before examining the results of each grade level in turn, it is useful to provide a general overview of the results. The tests developed for Grades One and Two seem superior to those developed for Grades Three and Four. There may be many reasons for this, but one hypothesis that should be kept in mind is that general listening skills may develop rapidly in the early grades after which development is tied more closely to verbal comprehension. Readers examining the test materials will notice that the tasks presented over the different grades require students to listen and recall. Essentially this appears to involve attention and memory. Difficult items in Grade Four tend to be items that require the recall of specific details. On the surface at least, such items would likely be as difficult for adults as for fourth graders. It may be the case that this kind of raw listening skill is reasonably well developed by Grade Three or Four, and that assessments beyond this should be tied more closely to comprehension tasks like extrapolation, inferencing, and relating ideas recalled from different parts of the text. In any event the data provided below attest to the success of the test developers in Grades One and Two.

For all of the grades, items are in a three alternative multiple choice format. In the analysis presented below, item difficulties between .40 and .89 are considered to be acceptable. Random guessing would yield difficulties of .33, and so .40 was considered sufficiently greater than random to be useful. It was felt that items with difficulties above .9 were not contributing to the information on student achievement. Discriminations of less than .3 were used to identify items that seemed to be measuring something beyond the domain of the remaining items.

Grade One. The Grade One test consists of 22 multiple choice items based on four passages. The first passage consists of a student describing a recipe. There are four items for this passage. All of the items have acceptable difficulties and discriminations (Tables 1 and 2). As a scale, the internal consistency is only .29, indicating that it would be inappropriate to use it on its own. The second passage consists of a teacher talking to her students. There are six items for this passage, and all have good discriminations and difficulties. Of all of the passages in Grade One, this has the highest internal consistency. Passage three is a story. One of the six items has a difficulty of less than .4 suggesting a reexamination of the item. In the final passage students listen to a conversation and respond to six items. As a group the internal consistency is -.12 indicating that there are some problems in the set. One item has a difficulty level of only .3, and on another item the discrimination is almost zero.

In spite of the three problem items, the test appears to be operating reasonably well overall. The test-retest reliability is .63 (correlation over time). When this value is adjusted for differences between means it is still a

Table 1

Grade One Item and Scale Results

Scale	Item	Alternative a	b	c	r(scale)	r(test) biserial	Easy	Hard	Poor disc.
1	1	.19	.71*	.08	.57	.55			
	2	.12	.16	.70*	.49	.46			
	3	.42*	.15	.41	.62	.51			
	4	.25	.15	.63*	.57	.33			
2	1	.49*	.33	.16	.49	.49			
	2	.22	.48*	.28	.43	.32			
	3	.31	.48*	.15	.53	.46			
	4	.10	.64*	.24	.49	.45			
	5	.61*	.12	.25	.54	.50			
	6	.11	.10	.76*	.49	.55			
3	1	.13	.17	.69*	.43	.39			
	2	.15	.62*	.22	.46	.40			
	3	.61*	.12	.25	.50	.42			
	4	.86*	.10	.02	.38	.54			
	5	.67*	.10	.20	.51	.55			
	6	.26	.38*	.33	.48	.33		x	
4	1	.49*	.16	.27	.32	.25			
	2	.34	.16	.43*	.39	.23			
	3	.13	.26	.60*	.39	.04			x
	4	.61*	.21	.16	.42	.29			
	5	.06	.71*	.19	.48	.45			
	6	.48	.30*	.16	.34	.28		x	

Table 2

Grade One Listening Test Scale Statistics

Scale	Items	Alpha	Mean	s.d.	Correlations				
					1	2	3	4	Total
1	4	.29	2.4	1.1	1.0	.3	.3	.1	.6
2	6	.39	3.5	1.4	.3	1.0	.3	.1	.7
3	6	.26	3.8	1.3	.3	.3	1.0	.2	.7
4	6	-.12	3.1	1.1	.1	.1	.2	1.0	.5
Total	**22**	**.53**	**12.8**	**3.2**	**.6**	**.7**	**.7**	**.5**	**1.0**

Test-Retest Reliability (N=29)

Time	Mean	s.d.	Correlation	Adjusted
1	10.6	3.6	.63	.47
2	12.8	3.0		

respectable .47. Differences between means over time can be considered as part of measurement error, or they can be disregarded from consideration if only one administration is foreseen. In this example, if we want to know how well the test orders children, then the correlation of .63 indicates that children are ranked in much the same way on two occasions. However, if we want to know whether the pupils get exactly the same scores on two occasions, then we need to allow the differences between means to influence the reliability. For Grade One students, for whom a listening test is a fairly novel task, the reliability of .47 suggests that for use as a screening device, the test should be readministered to any child that seems at risk on the first administration.

The correlation among the scales (passages) are not high, suggesting that there may be factors other than a general listening ability influencing the performance. Differential motivational patterns might produce such an effect. A principal components analysis was carried out on tetrachoric correlations among items. Four factors were rotated in an attempt to see if the item structure reflected the passage structure. Factor loadings of .3 or greater are shown in the Tables (Table 3). The first two components were made up of a mixture of almost all of the items in passages 1, 2 and 3. Item 4 in passage 1 did not load on any of the factors. Item 4 on passage 2 loaded on factor 3. Apart from these anomalies, the item relationships suggested a coherent structure.

Items from the final passage behaved strangely. The first two loaded in a bipolar fashion on factor 4, and the final four had bipolar loadings on factor 3. All of the bipolarities related to mood or intonation, and the results suggested that the students did not seem to respond to changing mood, so that if they were able to correctly identify cheerful intonation in one question, they didn't do well on "commanding" intonation on the next question. Two of the

Table 3

Grade One Item Factor Analysis

Scale	Item	Communality	I	II	III	IV
1	1	.48	.59			
	2	.39	.61			
	3	.41		.57		
	4	.06				
2	1	.28	.42			
	2	.43	.58			
	3	.30	.52			
	4	.31			.34	
	5	.29		.41		
	6	.37	.45		.30	
3	1	.15	.31			
	2	.19	.36			
	3	.39		.62		
	4	.32	.51			
	5	.29		.48		
	6	.46		.52		.32
4	1	.69				.81
	2	.60				-.74
	3	.47			-.66	
	4	.80		.43	-.77	
	5	.49		.50	.43	
	6	.58			.73	

items from the passage were previously shown to have problems. Perhaps by dispersing the items any problems that arise from adjacency would be overcome.

The norms provided for the test scores indicate that there is good individual discrimination. Variation among the class means also shows that the test could be useful for general class monitoring (Tables 4 and 5). Lower and upper values in Table 5 are 95% confidence intervals around class means.

Table 4

Norms for Grade One Listening

Score	Percentile
4	1
5	1
6	2
7	4
8	7
9	14
10	26
11	35
12	45
13	57
14	68
15	77
16	86
17	94
18	97
19	99
20	99
21	99
22	99

Table 5

School/Class Scores for Grade One Listening

School	Class	Mean	Std. Dev.	Lower	Upper
A	A	10.2	2.1	8.6	11.8
	B	11.3	4.1	9.1	13.5
	Total	**10.7**	**3.7**	**9.48**	**12.0**
B	A	13.6	3.3	12.3	14.3
	B	15.0	2.1	14.1	15.8
	C	13.5	3.3	12.3	14.8
	Total	**14.0**	**2.9**	**13.3**	**14.0**
C	A	12.0	2.5	10.9	13.1
	B	13.4	2.5	12.2	14.6
	Total	**12.7**	**2.6**	**11.8**	**13.5**
D	A	13.1	3.0	11.9	14.4
	B	13.1	3.2	11.5	14.7
	C	15.6	3.0	14.2	17.1
	D	13.2	3.5	11.7	14.7
	Total	**13.7**	**3.3**	**13.0**	**14.5**
E	A	12.2	2.8	11.1	13.4
F	A	11.8	3.2	10.3	13.3
	B	13.7	3.0	12.5	14.9
	Total	**12.9**	**3.2**	**11.9**	**13.9**
City		13.1	3.2	12.7	13.5

Grade Two. There are 21 items on the Grade Two listening test. These are based on four passages. The first passage is a tape of a teacher singing a song. There are six items of which the first five have acceptable difficulty and discrimination indices (Tables 6 and 7). The biserial correlation between the sixth item and total test score is only .19, which indicates that this item does not seem to be tapping the same skills as the remainder of the test. As a group of 6 the homogeneity of the items is .50 which is high for a short test.

In the second passage, the students listen to a poem and respond to 6 items. Three of the six items have difficulty levels of .90 or greater. These items should be reexamined to ensure that no cues are given in the questions themselves. Perhaps as a result of the three items, the internal consistency of the items is fairly low (.35).

There are four items from the third passage which is a student interview. All of these have acceptable difficulty levels, although for three of them, the difficulty levels are greater than 80%. This leads to a high average (3.1 out of 4) for this part of the test.

The final passage consists of an adult describing five series of three pictures each. There are five questions and each has acceptable difficulty and discrimination. As in the case of the previous passage the internal consistency is less than .4, but this is not uncommon for short subscales.

None of the passage scores has a correlation with another that exceeds .3. This is partly due to the shortness of the scales, but it may again reflect differential listening patterns. Some students may find certain passages more interesting than others and their attention may wander during uninteresting passages.

Table 6

Grade Two Item and Scale Results

Scale	Item	Alternative			r(scale)	r(test)	Easy	Hard	Poor
		a	b	c		biserial			disc.
1	1	.68*	.11	.21	.48	.63			
	2	.19	.77*	.04	.64	.72			
	3	.04	.76*	.20	.49	.40			
	4	.62*	.17	.21	.63	.50			
	5	.05	.11	.83*	.62	.78			
	6	.15	.21	.64*	.39	.19			x
2	1	.29	.49*	.22	.53	.34			
	2	.18	.68*	.14	.55	.47			
	3	.02	.96*	.02	.36	.47	x		
	4	.04	.90*	.06	.47	.52	x		
	5	.06	.03	.90*	.47	.46	x		
	6	.69*	.13	.18	.55	.55			
3	1	.07	.11	.82*	.55	.50			
	2	.81*	.14	.05	.58	.46			
	3	.13	.84*	04	.53	.59			
	4	.64*	.09	.27	.65	.36			
4	1	.25	.64*	.09	.59	.46			
	2	.68*	.25	.05	.59	.42			
	3	.06	.13	.79*	.52	.56			
	4	.78*	.07	.14	.53	.46			
	5	.88*	.11	.01	.42	.52			

Table 7

Grade Two Listening Test Scale Statistics

Scale	Items	Alpha	Mean	s.d.	Correlations				
					1	2	3	4	Total
1	6	.50	4.3	1.4	1.0	.2	.2	.2	.7
2	6	.35	4.6	1.1	.2	1.0	.2	.3	.7
3	4	.33	3.1	.9	.2	.2	1.0	.2	.6
4	5	.38	3.8	1.1	.2	.2	.2	1.0	.6
Total	**21**	**.63**	**15.8**	**3.0**	**.7**	**.7**	**.6**	**.6**	**1.0**

Test-Retest Reliability (N=30)

Time	Mean	s.d.	Correlation	Adjusted
1	16.4	2.7	.76	.62
2	17.8	2.1		

As a total test, the internal consistency was .63. Test-retest reliability was a strong .76, and even after adjustment for the differences in means the value was a respectable .62.

A four factor solution was chosen to represent the structure of the test. The first factor is clearly defined by items from passage 4 (Table 8). The second factor is identified with items from the first passage where four out of the six items have loadings of .5 or greater. Factors 8 and 4 are a mixture of items from the second and third passages. The overall impression is that the factor structure is very much based on the passage divisions of the test.

The test norms show that the test scores are skewed downward indicating that the test has a low ceiling (Table 9). The median score is about 16 out of 21 and few students have scores below 10. There is reasonable discrimination among classes given the low ceiling (Table 10). The Listening test would appear to be most useful for identifying students who are having difficulty in the area. For most students the challenge of the test is not overwhelming.

Table 8

Grade Two Item Factor Analysis

Scale	Item	Communality	I	II	III	IV
1	1	.51	.55		.35	
	2	.72		.79		
	3	.46		.50		
	4	.64		.79		
	5	.78		.79		
	6	.24			-.44	
2	1	.42				.58
	2	.19	.37			
	3	.77	.34		.76	
	4	.63				.78
	5	.66			.54	.60
	6	.45			.56	
3	1	.24	.32			
	2	.23				.39
	3	.42				.52
	4	.48			.62	
4	1	.44	.59			
	2	.38	.58			
	3	.43	.59			
	4	.42	.61			
	5	.27	.40			

Table 9

Norms for Grade Two Listening

Score	Percentile
6	1
7	2
8	3
9	4
10	5
11	7
12	10
13	17
14	26
15	35
16	46
17	60
18	75
19	87
20	85
21	89

Table 10

School/Class Scores for Grade Two Listening

School	Class	Mean	Std. Dev.	Lower	Upper
A	A	16.0	2.5	13.9	18.1
	B	15.3	2.5	14.3	16.3
	C	17.7	2.3	16.4	19.0
	Total	**16.1**	**2.6**	**15.3**	**16.9**
B	A	17.3	2.1	16.4	18.1
	B	17.1	2.1	16.3	17.9
	C	15.6	3.1	14.4	16.8
	Total	**16.6**	**2.6**	**16.0**	**17.2**
C	A	15.3	3.3	13.9	16.7
	B	16.4	3.6	14.6	18.1
	Total	**15.8**	**3.4**	**14.7**	**16.9**
D	A	15.0	2.4	13.6	16.4
	B	16.3	2.0	16.0	17.6
	C	15.4	2.9	14.3	16.6
	Total	**15.8**	**2.6**	**15.3**	**16.6**
E	A	18.0	2.1	17.1	18.9
F	A	13.3	3.7	11.8	14.8
	B	16.0	2.8	14.8	17.2
	Total	**14.6**	**3.5**	**13.6**	**15.6**
City		16.1	2.9	15.7	16.4
City		13.1	3.2	12.7	13.5

Grade Three. There are 22 items in the Grade Three test. The first four are based on a passage of student talk entitled, "Wonder Soup." The second passage is a story, and there are 12 items that elicit information from the story. The final six items are based on a poem. Of the 22 items, 18 have appropriate difficulties and discrimination indices (Tables 11 and 12). Ninety-seven percent of the students answered the first item correctly. This item asked the students to choose from among three titles. Since it occurs right after they have heard the story its ease may be accentuated. All of the other items for that passage were within tolerance, although as a group they were very easy with the average score for the four items being 3.4.

The second passage was by far the longest, and so it is not surprising that the three other problematic items were from this group. Two items had difficulty levels less than .4. Both these (items 1 and 7) were questions seeking the main idea. In both cases, the creative student might have chosen a different alternative. The final item from this passage (item 12) had an appropriate difficulty level, but its correlation with the total test score was only .25. As a group, the 12 items were reasonably homogeneous (alpha = .50). The final passage (a poem) had 6 items that were very homogeneous. All had acceptable difficulties and discriminations.

For the test as a whole, the internal consistency was .61 indicating that overall the items seemed to be measuring a common attribute, although passage intercorrelations were never greater than .3. The test-retest correlation was .60. This indicated that the test ordered students in about the same way on two separate occasicns. Unfortunately, when the reliability was corrected for the differences between the pre and post test means, the value dropped to -.01. This resulted from a dramatic upward shift in the means (from 14.2 to 18.0) for the 30 students involved in the reliability study. Further study should be

Table 11

Grade Three Item and Scale Results

Scale	Item	Alternative a	b	c	r(scale)	r(test) biserial	Easy	Hard	Poor disc.
1	1	.00	.97*	.02	.35	.59	x		
	2	.07	.83*	.09	.60	.60			
	3	.75*	.07	.19	.68	.34			
	4	.06	.04	.88*	.56	.49			
2	1	.16	.27*	.57	.38	.47		x	
	2	.60*	.31	.09	.39	.40			
	3	.04	.30	.63*	.41	.48			
	4	.09	.34	.57*	.47	.54			
	5	.43	.40*	.16	.42	.38			
	6	.41	.49*	.10	.41	.44			
	7	.30	.31	.37*	.49	.62		x	
	8	.48*	.11	.38	.31	.32			
	9	.35	.53*	.12	.38	.41			
	10	.47	.50*	.03	.51	.57			
	11	.17	.47*	.35	.38	.38			
	12	.29	.14	.59*	.17	.25			x
3	1	.34	.62*	.01	.61	.50			
	2	.26	.04	.70*	.48	.35			
	3	.08	.77*	.14	.41	.26			
	4	.10	.24	.64*	.60	.48			
	5	.90*	.04	.04	.52	.63			
	6	.73*	.13	.10	.47	.49			

Table 12

Grade Three Listening Test Scale Statistics

Scale	Items	Alpha	Mean	s.d.	Correlations			
					1	2	3	Total
1	4	.26	3.4	.8	1.0	.3	.2	.5
2	12	.50	5.9	2.3	.3	1.0	.3	.9
3	6	.44	4.4	1.3	.2	.3	1.0	.6
Total	**22**	**.61**	**13.7**	**3.3**	**.5**	**.9**	**.6**	**1.0**

Test-Retest Reliability (N=30)

Time	Mean	s.d.	Correlation	Adjusted
1	14.2	3.1	.60	-.01
2	18.0	1.7		

undertaken to see if this is an anomaly, or if it is a common occurrence. The district mean was 14.1 suggesting that the pretest value was at an appropriate level.

A three factor solution was tried for the factor analysis (Table 13). A fairly strong first factor (20% of the total variance) occurring in the unrotated solution corroborated the common thread indicated by the alpha coefficient. Eleven of 22 items had correlations of .4 or greater with the first principal component. After rotating three factors, no pattern emerged. Two items had communalities less than .1. These were item 3 in passage 1, and item 8 in passage 2. Low communalities suggest that the items have a weak relationship with the other items. Four of the six items from the third passage loaded on the second factor. Items from the second passage loaded mostly on factors 1 and 3, while those from the first passage loaded on all factors.

The median test score for the district was about 14 out of 22. The test showed good potential for discriminating among classes, and among students especially those at the lower end of the continuum (Tables 14 and 15). Before adoption however, the problems associated with substantial increases in test scores should be resolved.

Table 13

Grade Three Item Factor Analysis

Scale	Item	Communality	I	II	III
1	1	.97	.64	.74	
	2	.48			.66
	3	.09			
	4	.28			.44
2	1	.27	.48		
	2	.20			.43
	3	.37			.56
	4	.37	.40		.37
	5	.42	.48	-.38	
	6	.29			.42
	7	.78	.77	.41	
	8	.06			
	9	.24			.47
	10	.74	.82		
	11	.37			.58
	12	.35		.43	
3	1	.22	.44		
	2	.11	.33		
	3	.12		.34	
	4	.47		.64	
	5	.65		.74	
	6	.49		.47	.51

Table 14

Norms for Grade Three Listening

Score	Percentile
5	1
6	2
7	3
8	6
9	10
10	13
11	20
12	30
13	41
14	51
15	63
16	75
17	84
18	91
19	95
20	99

Table 15

School/Class Scores for Grade Three Listening

School	Classes	Mean	Std. Dev.	Lower	Upper
A	A	13.7	3.1	12.5	14.8
	B	13.8	2.8	12.1	15.4
	Total	**13.7**	**3.0**	**12.7**	**14.6**
B	A	14.3	3.2	13.1	15.5
	B	13.7	2.2	12.9	14.5
	Total	**14.0**	**2.7**	**13.3**	**14.7**
C	A	13.1	3.4	11.7	14.5
	B	14.5	2.0	13.1	15.9
	Total	**13.4**	**3.2**	**12.3**	**14.6**
D	A	16.3	2.8	15.7	17.9
	B	13.2	3.2	12.0	14.4
	C	13.4	2.6	11.8	14.9
	Total	**14.7**	**3.4**	**13.8**	**15.5**
E	A	13.9	3.4	12.6	15.2
F	A	14.6	2.6	13.5	15.7
	B	13.8	3.1	12.6	15.1
	Total	**14.2**	**2.9**	**13.4**	**15.1**
City		14.1	3.1	13.7	14.5

Grade Four. The Grade Four version of the test met with mixed success. Five of 30 items had difficulties of .9 or greater and two had difficulties of less than .4. In addition, one item had a test biserial of less than .3. Twenty-two items had difficulty and discrimination indices that were acceptable (Tables 16 and 17).

The test was composed of six passages. The first was a short story on which 6 items were based. Two of the items were too easy, with 99% and 94% of the students answering the first and third items correctly. The second passage was a poem with 4 items. Two of the four items had difficulties greater than .9. The third task required the students to trace a pattern according to instructions. It had very appropriate difficulty, but there may be a "learning how to do the test" factor influencing performance, and it would have been better to have a few more items of the same format.

There were 7 items for the News Report and 6 items for the Weather Report. While one item from each of these passages was difficult, in general these questions were well designed. The final passage was a conversation. There were 6 items of which one had a difficulty that exceeded .9.

The item homogeneity for the test as a whole was only .58 (fairly low for a 30 item test), however the test-retest correlation was excellent (.70). The problem encountered in Grade 3 did not occur with this group, since the adjusted reliability was a respectable .62. The passage scores did not correlate well with each other. This confirmed the observation made about item homogeneity and seems consistent with the idea noted at the beginning of this document that by the time students reach grade 4, listening tests may be measuring attention more than listening skill.

The factor analysis was based on Pearson correlations rather than tetrachorics because many of the tetrachorics exceeded one. The factor

Table 16

Grade Four Item and Scale Results

Scale	Item	Alternative a	 b	 c	r(scale)	r(test) biserial	Easy	Hard	Poor disc.
1	1	.99*	01	.00	.33	.71	x		
	2	.29	.00	.71*	.48	.39			
	3	.03	.34*	.02	.28	.40	x		
	4	.04	.22	.75*	.51	.29			
	5	.08	.82*	.09	.53	.42			
	6	.65*	.20	.14	.58	.30			
2	1	.34*	.00	.05	.42	.26	x		
	2	.88	.03	.03	.64	.41			
	3	.72*	.07	.20	.76	.46			
	4	.03	.00	.97*	.26	.36	x		
3	1	.27	.59*	.06	1.00	.35			
4	1	.82*	.35	.03	.47	.48			
	2	.05	.38*	.57	.42	.03		x	x
	3	.34	.44*	.22	.38	.21			x
	4	.33	.18	.48*	.45	.30			
	5	.61*	.14	.24	.46	.53			
	6	.03	.28	.89*	.45	.38			
	7	.63*	.05	.32	.34	.29			
5	1	.21	.62*	.17	.50	.48			
	2	.21	.37	.42*	.55	.51			
	3	.64*	.22	.13	.54	.42			
	4	.09	.08	.82*	.47	.42			
	5	.16*	.40	.44	.34	.38		x	
	6	.29	.27	.44*	.54	.40			
6	1	.11	.03	.85*	.42	.46			
	2	.01	.74*	.25	.48	.28			
	3	.53*	.18	.26	.47	.39			
	4	.94*	.03	.02	.38	.53	x		
	5	.63*	.31	.06	.52	.37			
	6	.15	.71*	.12	.52	.53			

Table 17

Grade Four Listening Test Scale Statistics

Scale	Items	Alpha	Mean	s.d.	Correlations						
					1	2	3	4	5	6	Total
1	6	.24	4.9	1.0	1.0	.1	.0	.2	.2	.1	.5
2	4	.23	3.5	.7	.1	1.0	.1	.1	.2	.2	.6
3	1		.6	.4	.0	.1	1.0	.1	.0	.2	.5
4	7	.24	3.8	1.4	.2	.1	.1	1.0	.2	.2	.7
5	6	.38	3.1	1.3	.2	.2	.0	.2	1.0	.3	.7
6	6	.28	4.4	1.2	.1	.2	.2	.2	.3	1.0	.6
Total	30	.58	20.3	3.5	.5	.3	.5	.7	.7	.6	1.0

Test-Retest Reliability (N=30)

Time	Mean	s.d.	Correlation	Adjusted
1	20.8	3.6	.70	.62
2	22.5	3.7		

structure was diverse (Table 18). Five factors were chosen for rotation although the eigenvalues provided ambiguous support for this. The first factor could be interpreted as a Weather Forecast factor with 4 of 6 items from this passage loading here. However, 4 items from other passages also loaded on this factor. The remainder of the factors were composed of items from several passages. An attempt was made to interpret the factors according to the skills specified for each item found in the administration manual. This attempt failed.

For a thirty item test, the instrument had good discrimination especially at the lower end (Table 19). The first quartile is a little high (18) but the third quartile of 23 leaves plenty of ceiling for high achievers. Perhaps the best use to be made of the test is to select children who are performing poorly for further diagnosis.

Ten of twelve classes had means within one point of the district average (Table 20). This suggests that either listening skills are well developed by Grade Four or the test lacks group discriminability. The preference here is for the first explanation. Careful examination of the test itself shows that the passages tap a wide variety of school related listening skills, from following instructions to interpreting and extending stories. Some of the passages are taken from ordinary life events, and have a face validity suitable for adults.

Table 18

Grade Four Item Factor Analysis

Scale	Item	Communality	I	II	III	IV	V
1	1	.51		.67			
	2	.22			.45		
	3	.30		.50			
	4	.38				-.50	
	5	.19			.41		
	6	.32			.40		
2	1	.51	.41				-.58
	2	.36		.45			
	3	.30		.45			
	4	.22				.46	
3	1	.30				.52	
4	1	.30			.51		
	2	.33					.56
	3	.40			.47		
	4	.27	.34			-.34	
	5	.26	.38				
	6	.28					.33
	7	.12					
5	1	.22	.34		.31		
	2	.30	.48				
	3	.30	.31			-.37	
	4	.26	.50				
	5	.31			.48		
	6	.12					
6	1	.23				.34	
	2	.16					.34
	3	.13					
	4	.22					.33
	5	.30		.34			
	6	.46	.46				.45

Table 19

Norms for Grade Four Listening

Score	Percentile
10	0
11	1
12	2
13	3
14	5
15	8
16	13
17	17
18	24
19	34
20	44
21	55
22	68
23	73
24	86
25	91
26	95
37	98
38	99
29	99
30	99

Table 20

School/Class Scores for Grade Four

School	Classes	Mean	Std. Dev.	Lower	Upper
A	A	20.6	4.2	18.9	22.2
B	A	20.1	2.8	13.0	21.1
	B	21.7	2.8	20.6	22.7
	Total	**20.8**	**2.9**	**20.1**	**21.6**
C	A	20.6	3.3	19.1	22.2
	B	20.0	3.6	18.0	22.0
	Total	20.4	3.8	**13.2**	**21.7**
D	A	19.8	3.6	17.6	21.6
	B	20.5	2.9	19.4	21.7
	C	20.0	3.2	18.7	21.2
	Total	**20.2**	**3.1**	**19.4**	**20.8**
E	A	22.3	3.2	21.1	24.7
	B	19.7	3.8	18.1	21.3
	Total	**20.3**	**3.3**	**18.6**	**22.2**
F	A	20.0	3.5	18.7	21.3
	B	21.5	3.8	13.5	23.7
	Total	**20.4**	**3.6**	**13.3**	**21.6**
City		20.5	3.5	20.1	21.0

Problem Two: Listening Assessment, Grades Five and Six, Discussion

As noted earlier, three problems concerning the area of listening were presented for investigation in the study. The second problem was: Do the conceptual domains of listening skills which underlie and form the basis of the construction of the Edmonton Public Listening Tests have empirical validity?

The following Action Plan and Timeline for this problem were included in the Proposal.

(1) June 84	Acquire and administer Edmonton Public Listening Tests to Grades Five and Six.
(2) July/August 84	Conduct statistical analysis of reliability and factor analysis of Edmonton Public Listening Tests; develop norms for Grande Prairie and compare to Edmonton norms.

In general, the Action Plan and Timeline were reasonably well followed. Battery A of the Edmonton Public Listening Tests, also known as the Tests of Listening Comprehension, was administered in June, 1983, and again in June, 1984. Statistical analyses were conducted by Dr. Tom Maguire; his findings follow this discussion.

The Listening Comprehension Tests were originally developed by Andrew Wilkinson, Leslie Stratta, and Peter Dudley in England and were published in 1976. Three separate test batteries were produced. Battery A is intended for 10-11 year old students (Grades 5, 6); Battery B for 13-14 year old students (Grades 8, 9) and Battery C for 17-18 year old students (Grade 11, 12, 13+).

The three batteries are basically the same, although they differ in detail. Each consists of spoken material recorded on audio tape along with questions relating to that material. Consumable Student Test Booklets are available for student responses. Multiple choice answer options are provided; however, the number of choices is not necessarily consistent from subtest to subtest, nor is the format of the answer options. Included in the Teacher's Manual are descriptions of the test content and type, marking and scoring procedures, and test statistics.

Each battery consists of a number of subtests designed to measure various elements of listening comprehension. These are:

> Tests of Content, designed to measure the ability to follow and understand a piece of informal exposition;
>
> Tests of Prediction, designed to measure the ability to infer missing parts of a conversation from what is actually heard or the ability to predict whether certain utterances are appropriate in a given context;
>
> Tests of Phonology, designed to measure the ability to understand differences in meaning brought about by varying intonation and stress;
>
> Tests of Register, designed to measure the ability to detect changes in the appropriateness of the spoken language used;
>
> Tests of Relationship, designed to measure the ability to detect the kinds of relationships existing between people from the language they employ.

In 1978 the Edmonton Public School Board #7, with the permission of the authors, revised the original tests so that they would be suitable for use in Edmonton Public Schools. Once the revisions were completed, the tests were field tested in Edmonton Public Schools during 1979. Later in 1979 final drafts were prepared and norms were calculated during the Spring of 1980. These are now included in the Teacher's Manual, entitled Administration Manual.

According to the authors of the Edmonton Public Listening Tests, as reported in the Administration Manual, Listening Comprehension Tests, Canadian Version (1980), the tests have a number of possible uses. They can be applied diagnostically at various stages in a student's school career to help teachers measure problems and levels of understanding in relation to other language abilities; they can be used as a test of understanding for speakers whose native language is not English; and they can be used as developmental materials, because the tests contain a large variety of samples of English which would be appropriate for exploration of facets and features of spoken language.

Problem Two--Listening Assessment, Grades Five and Six, Findings

The results of the statistical analyses reported by Dr. Maguire reveal that the answer to Problem Two is no. There was a lack of strong relationship among the items on the Edmonton Public Listening Tests, and little evidence that the items clustered in the alignment suggested by the subtests. Thus, it appeared that the traits measured by the subtests were not independent of each other, and that the conceptual domains of listening skills which form the basis of the construction of the Edmonton Public Listening Tests do not have empirical validity.

Factor Analysis of the Edmonton Public Listening Tests

(Maguire, 1984)

The seventy-five items of the listening tests were correlated with each other, and a principal components analysis was carried out. About one third of the items had difficulty levels that exceeded .90, and another 17 had difficulties between .8 and .9. Consequently, the correlations among the items were not very large. In the analysis itself, there were 30 roots greater than one, which is a further indication of the lack of strong relationship among items.

In an attempt to see how well the structure of the components matched the structure suggested by the subtests, an orthogonal Procrustes rotation was carried out. In this rotation, a target matrix consisting of ones and zeros is set up so that it matches the assignment of items to subtests. An orthogonal least squares fit is then carried out in an attempt to match the target. In spite of this effort there is little evidence that the items cluster in the alignment suggested by the subtests.

The intercorrelations among subtests were similar to those found in the Administration Manual, falling in the vicinity of .3 to .4. This suggests that the traits measured by the subtests are not independent of each other, and so it is little wonder that the principal components analysis was unable to satisfy an orthogonal relationship among the items.

Table 22 provides Grades Five and Six percentile norms. Table 23 provides school/grade score comparisons at the Grades Five and Six levels.

Table 21

Correlations Among Subtests for Grades Five and Six

Sample						
Content 1	1.00					
Content 2	.41	1.00				
Prediction	.39	.38	1.00			
Phonology	.26	.26	.32	1.00		
Register	.38	.37	.38	.28	1.00	
Relationship	.46	.36	.34	.28	.40	1.00

Table 22

Grades Five and Six Percentile Norms for Listening

Score	Grade 5	Grade 6
37	1	
38	2	
39	2	
40	2	
41	2	
43	4	
44	6	1
45	8	2
46	9	2
47	11	3
48	13	3
49	14	5
50	17	6
51	21	8
52	25	11
53	31	14
54	38	17
55	43	21
56	49	24
57	56	28
58	63	32
59	69	38
60	75	43
61	82	50
62	86	58
63	87	67
64	92	75
65	95	81
66	97	86
67	98	92
68	99	96
69	99	98
70	99	99

Table 23

School/Grade Scores for Grade Five Listening

School	Teacher	Mean	Std. Dev.	Lower	Upper
A	1	52.7	8.6	49.3	56.1
	2	57.1	6.0	54.9	59.4
	Total	**55.0**	**7.6**	**52.9**	**57.1**
B	1	59.6	5.5	57.6	61.7
	2	56.9	4.8	55.1	58.7
	Total	**58.3**	**5.3**	**56.9**	**59.7**
C	1	53.7	6.1	51.2	56.1
	2	55.7	5.6	53.5	57.8
	3	54.1	5.7	51.8	56.5
	Total	**54.5**	**5.8**	**53.2**	**55.8**
D	1	53.2	6.8	50.8	55.7
E	1	57.0	5.7	54.1	59.9
	2	54.8	6.9	52.3	57.3
	Total	**55.6**	**6.5**	**53.6**	**57.5**
City		55.4	6.5	53.6	56.3

Grade 6 Listening Test

School	Teacher	Mean	Std. Dev.	Lower	Upper
A	1	59.6	6.1	57.4	61.8
	2	61.7	6.1	59.5	64.0
	Total	**60.7**	**6.2**	**59.1**	**62.3**
B	1	59.9	5.6	57.5	62.2
	2	61.4	4.1	59.7	63.0
	Total	**60.6**	**4.9**	**59.2**	**62.1**
C	1	59.0	5.6	56.9	61.1
	2	55.6	6.7	53.0	58.2
	3	58.9	6.0	56.7	61.1
	Total	**57.9**	**6.2**	**56.5**	**59.2**
D	1	61.1	5.7	58.6	63.6
E	1	61.0	4.7	59.2	62.7
	2	62.6	3.8	60.0	65.1
	Total	**61.3**	**4.5**	**59.9**	**62.8**
City		59.9	5.8	59.1	60.6

Problem Three: Listening Assessment, Correlations Among Language Variables, Discussion

As noted earlier, three problems concerning the area of listening were presented for investigation in the study. The third problem was: What other language arts skills correlate most highly with listening skills?

The following Action Plan and Timeline for this problem were included in the Proposal.

July/August 85	Conduct correlation studies of listening achievement, reading achievement, I.Q., and achievement in written composition.

In general, the Action Plan and Timeline were reasonably well followed. Correlation studies were carried out at each of Grades Two to Six, and the results reports in Dr. Maguire's 1986 document.

It should be noted that correlations among the language arts variables differed from grade to grade in terms of available data. For example, at Grade Two listening scores were correlated only with reading scores, while at Grade Six correlation variables included, in addition to listening, reading, writing, spelling, verbal ability, quantitative ability, and non-verbal ability.

Problem Three: Listening Assessment, Correlation Among Language Variables, Findings

The results of the statistical analyses reported by Dr. Maguire reveal that the answer to Problem Three differs by grade level. At Grade Two, the listening scores correlated well with the decoding and comprehension scores of the reading test. At Grade Three, these correlations were lower, and there was little correlation between listening and spelling scores. At Grade Four, the listening scores correlated well with the decoding and comprehension scores of the reading test and with the verbal I.Q. scores but poorly with the general impression (writing) and spelling scores. At Grade Five, the correlations were highest between the listening, reading comprehension and verbal ability scores. At Grade Six all of the correlations exceeded .4, suggesting that there is a fairly strong language ability running through the test at this level.

Analysis of Relationships Among Language Arts Scores

(Maguire, 1986)

Grade Two. Correlations were calculated between total listening scores obtained on the GPLT and reading scores measured several months earlier. The correlations are shown in Table 24, and indicate a general relationship between listening and reading of about .4. This is an acceptable value. If it were much higher, say .6, there would be concern that the listening test is too much a comprehension test. If the value were lower, there would be fear that the test lacked validity. Listening should have some relationship to verbal comprehension, but it should be distinct from it.

Grade Three. Correlations of total listening scores obtained on the GPLT with reading skills were somewhat lower than at the second grade level (r=.24 and .38 for decoding and comprehension respectively). The correlation between this listening test (June, 1985) and an earlier version of the second grade listening test administered one year earlier was .35 which is a bit low but could be accounted for by the poorer psychometric qualities of the earlier test.

Grade Four. Correlations between total listening scores obtained on the GPLT and outside variables such as verbal ability are all in the neighborhood of .4, indicating that if the attention hypothesis is true, it may be related to verbal tasks. (See Instrument Evaluation Report, Introduction, pp. 36, 37.)

Table 24

Correlations With Reading Variables, Grade Two

Variable	Mean	S.d.	Correlations 1	 2	 3
Listening	16.3	2.9	1.00	.38	.42
Decoding	44.5	5.7	.38	1.00	.78
Comp.	38.2	8.2	.42	.78	1.00

Listening Test was administered June of Grade 2.
Decading and Comprehension were administered in June of Grade 1.

Table 25

Correlations With External Variables , Grade Three

Variable	Mean	S.d.	Correlations 1	 2	 3	 4	 5
Listening 1	15.5	2.2	1.00	.35	.17	.30	.16
Listening 2	14.3	2.9	.35	1.00	.24	.38	.12
Decoding	60.3	7.5	.17	.24	1.00	.78	.68
Comp.	53.3	11.7	.30	.38	.78	1.00	.58
Spelling	17.9	7.7	.16	.12	.68	.58	1.00

Listening 1 is a preliminary version of the Grade 2 Listening Test which was administered in October of Grade 3.
Listening 2 is the current Listening Test administered in June of Grade 3.
Decoding is a reading subtest administered in Grade 2.
Comp. is a Comprehension subtest administered in Grade 2.
Spelling is a Divisional Spelling test administered in Grade 2.

Table 26

Correlations With External Variables, Grade Four

Variable	Mean	s.d.	Correlations								
			1	2	3	4	5	6	7	8	9
Listening 1	17.5	2.7	1.00	.42	.42	.53	.27	.43	.44	.45	.39
Listening 2	20.5	3.4	.42	1.00	.43	.43	.27	.38	.49	.41	.33
Decoding	51.4	7.1	.42	.43	1.00	.82	.45	.70	.77	.55	.53
Comp.	58.0	12.9	.53	.43	.82	1.00	.38	.62	.78	.61	.56
G. Imp.	7.7	2.6	.27	.27	.45	.38	1.00	.47	.41	.36	.23
Spelling	24.7	9.0	.43	.38	.70	.62	.47	1.00	.65	.50	.39
VIO	68.4	15.3	.44	.49	.77	.78	.41	.65	1.00	.65	.58
Quant.	44.7	7.9	.45	.41	.55	.61	.36	.50	.65	1.00	.65
NVIO.	63.7	8.9	.39	.39	.53	.56	.23	.39	.58	.65	1.00

Listening 1 is a preliminary version of the Grade 3 Listening test administered in October of Grade 4.
Listening 2 is the current Listening test administered in June of Grade 4.
Decoding is a Reading subtest administered in Grade 3.
Comp. is the Comprehension subtest administered in Grade 3.
Gen. Impression is Writing General Impression administered in Grade 3.
Spelling is the Spelling test administered in June of Grade 3.
VIO is Verbal IQ percentiles.
Quant. is Quantitatives 1Q percentiles.
NVIO is non verbal IQ percentiles.

Grade Five. Correlations between total listening scores on the Edmonton Public Listening Tests and seven language arts scores were examined. These variables were reading (two scales--decoding and comprehension), writing (general impression), spelling, and three facets of general ability (verbal ability, quantitative ability, and non-verbal ability) as measured by the Lorge-Thorndike Test. The listening test had correlations of .5 or greater with reading comprehension and verbal ability, whereas with writing and spelling the values were less than three. In general, there did not seem to be a general language arts ability running through the tests at this level.

Grade Six. Here a similar set of scales to those of Grade Five were used, except that no writing skills were assessed. The correlations between listening and all of the other variables exceeded .4, and in the case of reading comprehension and verbal ability the value was .6. This suggests that there is a fairly strong language ability running through the battery at this level.

Table 27

Correlations With External Variables, Grade Five

Listening	1.00	.40	.55	.28	.22	.59	.38	.40
Decoding		1.00	.81	.46	.65	.72	.67	.60
Comprehension			1.00	.48	.46	.83	.67	.64
Writing				1.00	.39	.43	.38	.36
Spelling					1.00	.47	.33	.26
Verbal Ability						1.00	.71	.64
Quantitative Ability							1.00	.71
Non-Verbal Ability								1.00

Table 28

Correlations With External Variables, Grade Six

Listening	1.00	.54	.60	.42	.60	.46	.43
Decoding		1.00	.73	.62	.75	.62	.52
Comprehension			1.00	.53	.82	.70	.53
Spelling				1.00	.59	.46	.27
Verbal Ability					1.00	.67	.53
Quantitative Ability						1.00	.74
Non-Verbal Ability							1.00

RESEARCH CLUSTER TWO--WRITTEN COMPOSITION

The following problems concerning the area of written composition were presented for investigation in the study.

1. Can the analytic scoring techniques that have been applied at secondary school levels be used reliably for the rating of written compositions at Grades Three, Four, and Five?
2. How does analytic scoring compare to holistic scoring in Grade Three for each of the following dimensions: (a) inter-rater reliability; (b) cost effectiveness; and (c) teacher satisfaction (both in terms of those teachers involved and those who receive the information from scoring)?
3. Are there significant differences in achievement in written composition in Grades Four and Five depending on the mode of writing required?

One major and one secondary purpose of the study underlie the development of these questions. These are (1) an examination of the relative value of two techniques for the scoring of written composition; and (2) an investigation of achievement in written composition based on different writing modes.

The Proposal noted that definitions of "holistic" and "analytic" scoring vary in the literature. For this study, the developers of the Proposal therefore decided to use the definitions of these terms as provided by Cooper and Odell in Evaluating Writing (1977).

Cooper and Odell define "holistic" scoring as a general impression scheme for examining the quality of written compositions. They write:

> Holistic evaluation is a guided procedure for sorting or ranking written pieces. The rater takes a piece of writing and either (1) matches it with another piece in a graded series of pieces or (2) scores it for the prominence of certain features important to the kind of writing or (3) assigns it a letter or number grade. The placing, scoring, or grading occurs quickly,

> impressionistically, after the rater has practised the procedure with other raters. The rater does not make corrections or revisions in the paper. Holistic evaluation is usually guided by a holistic scoring guide which describes each feature and identifies high, middle and low quality levels for each feature. (p. 3)

Cooper and Odell then provide several types of holistic evaluations. One type is the "essay" or "general impression" type. It is defined as

> . . . a series of complete pieces arranged according to quality. At the end of the series is an exemplary piece, at the other an inadequate one. The pieces which make up the scale are usually selected from large numbers of pieces written by students like those with whom the scale will be used. A rater attempts to place a new piece of writing along the scale, matching it with the scale piece most like it. (p. 4)

These authors then define another type as "analytic."

> An analytic scale (a holistic evaluation device) . . . is a list of the prominent features or characteristics or writing in a particular mode. The list of features ordinarily ranges from four to ten or twelve, with each feature described in some detail and with high-mid-low points identified and described along a scoring line for each feature. (p. 7)

In this study, the terms "holistic/general impression" were used interchangeably to describe what Cooper and Odell labelled the essay type of scoring. The term "analytic" was not preceded by "holistic" in this study, even though it is in fact a holistic type of marking.

Problem One: Holistic/Analytic Scoring, Grades Three to Five, Discussion

The following Action Plan and Timeline for this problem were included in the Proposal.

(1) June 83	Develop a writing stimulus for Grade Three (a descriptive writing task as used in the pilot study, June, 1983).
(2) February 84	Develop two writing stimuli for Grades Four and Five, requiring writing of differing modes.

(3) May 84	Have students write the compositions required by the aforementioned stimuli (students in Grades Four and Five randomly assigned to differing modes).
(4) July/August 84	Have teacher marking teams assess the compositions analytically, and conduct statistical analyses of inter-rater reliability of results.

In general the Action Plan and Timeline were well followed.

In June, 1983, a pilot study was conducted by Grande Prairie School District #2357, during which a stimulus for descriptive writing was developed and administered to all Grade Three students. The compositions produced in this pilot study were marked holistically (general impression) by a team of Grade Three teachers. Inter-rater reliability was found to be high. It was decided to use a descriptive writing task again for the June, 1984, assessment at the Grade Three level. During the next several months, two writing stimuli were developed for Grades Four and Five, involving the use of a narrative task and a persuasive task.

Grade Three. In June, 1984, each student at the Grade Three level was required to produce a piece of descriptive writing. The stimulus was the same for all students. It involved a poster-sized color picture (50cm by 75cm) of nine hot air balloons in various stages of inflation. The background showed Bear Creek, trees, spectators, and houses. The students received the following directions:

> Look at the poster that your teacher will show you. This is a picture of a scene in Bear Creek Park here in Grande Prairie.
>
> Imagine that you are on a hill above the scene. Write about the scene so that someone who is not there can form a picture of it in his mind. Don't tell a story. Write a description. Paint a picture with your words. . . .

The pieces of writing were assigned a random identification number and scored by a team of four Grade Three teachers in July, 1984. These compositions were first scored holistically (general impression). They were then scored analytically for the following features (1) content; (2) development; (3) sentence structure; (4) vocabulary; and (5) writing conventions. The criteria were based on descriptors established for the June, 1984, Grade Six provincial language arts achievement testing in writing. Both the holistic and analytic scoring were done by the same marking team. The descriptors for both the holistic and analytic scoring are provided in the Grade Three Teacher Marking Package (November, 1984). This Package and the Teacher Marking Package for Grades Four/Five are appended at the end of this Report.

At this point it must be emphasized that both the Grades Three and Four/Five Teacher Marking Packages were developed specifically for inservice staff development with teachers at the Grades Three to Five levels in Grande Prairie School District #2357. These packages include student writing instructions, scoring descriptors, sample pieces of student writing, the scores given to each sample piece of writing by the marking team, and the results of the statistical analyses of the scoring.

This writing is reflective of the students in only this school jurisdiction, and is not necessarily reflective of student writing across the province. Thus, the Teacher Marking Packages should NOT be used as a basis of comparison of student writing for any purpose or at any grade level. They are provided along with this Report to indicate what was done locally by Grande Prairie School District #2357 in order to answer the research questions posed in this study. Further information on the assessment of writing at the elementary level may be obtained from the Student Evaluation Branch, Alberta Education.

Grades Four and Five. In June, 1984, each student at the Grades Four and Five levels was required to produce a piece of writing, involving either a narrative or a persuasive writing task. Two topics were provided, one concerned with camping and the other concerned with becoming teacher for a day. Students were permitted to select their topic but the teachers randomly assigned the writing task based on the selected topic.

Concerning the topic of camping, the students received the following instructions:

> Task A (Persuasive)
>
> If you were planning a camping trip, what types of things would you want to take with you?. . . Imagine that you would be camping for two days and could take only what you could put in the trunk of a car. Give reasons why you would take the things that you include.
>
> Task B (Narrative)
>
> Write a story that tells what actually happens on Jason's camping trip, or on a camping trip that you went on. The story does not have to be true. . .

Concerning the topic of teacher for a day, the instructions were:

> Task A (Persuasive)
>
> If you were elected teacher for a day, what kinds of activities would you plan for your classmates? Give good reasons for including the activities you have chosen.
>
> Task B (Narrative)
>
> Write a story that tells what happens on the day that Lori is "teacher for a day". . .

As with the Grade Three assessment, the pieces of writing were assigned a random identification number. They were then scored by a team of nine Grades Four and Five teachers. The papers were first scored holistically (general impression) and then analytically with the same features as those used by the Grade Three markers. For more information, and for the lists of descriptors, see the Teacher Marking Package, Grades Four/Five, November, 1984 appended at the end of this Report.

Problem One: Holistic/Analytic Scoring, Grades Three to Five, Findings

Statistical analyses of the pieces of student writing provided by the June, 1984 assessment were conducted by Dr. Tom Maguire. The Teacher's Marking Packages contained the Results Summaries which are reproduced here.

Problem One in Research Cluster One--Written Composition asked: Can the analytic scoring techniques that have been applied at secondary school levels be used reliably for the rating of written compositions at Grades Three, Four, and Five? The answer to this question is an unqualified yes at all grade levels. Inter-rater reliabilities at all grade levels were .78 or higher for the holistic/general impression scoring and for all features of the analytic scoring.

Grade Three. Table 29 provides the overall district means and standard deviations for both the holistic/general impression and analytic scores for Grade Three. Table 30 provides the correlations among the various features examined. Table 31 shows the percentile rankings for the various scores obtained by the students based on the holistic and analytic scoring.

Descriptive writing, the mode employed for this assessment, attempts to create a verbal portrait of a subject. According to Temple and Gillet (1984), "Descriptions can be challenging to write because the logic of description is not always easy. . . Description (is) a difficult form for children of all ages to master but a valuable one to attempt." (p. 219)

It is interesting to note that no mean score was under 2.4, and none was higher than 2.8. It may be assumed that, as children learn more about this mode and as they mature cognitively, their ability to write in this mode will increase. Nevertheless, at this level, these scores seem most appropriate. A score of 2.8 places students in the 50-70 percentile across the various analytic features. The correlations among the variables suggests that students who function well in any one variable are likely to function well in the others.

Table 29

Overall District Results for Grade Three Written Composition

	Mean	Std. Dev.	Confidence Int.
General Impression	2.6	.7	2.5 - 2.7
Content	2.8	.7	2.7 - 2.8
Vocabularly	2.6	.7	2.6 - 2.7
Conventions	2.6	.8	2.5 - 2.7
Development	2.4	.8	2.3 - 2.5
Sentence Structure	2.7	.8	2.6 - 2.8
Average of Analysis Components	2.6	.6	2.6 - 2.7

Table 30

Correlations Among Written Composition Variables, Grade Three

	Impr.	Cont.	Devl.	Sent.	Vocab.	Conv.	Averg.
General Impression	1.00	.72	.60	.68	.61	.68	.77
Content		1.00	.66	.64	.76	.62	.86
Development			1.00	.64	.65	.55	.83
Sentence				1.00	.67	.71	.87
Vocabularly					1.00	.58	.85
Conventions						1.00	.83
Average of Analysis Components							1.00

Table 31

Percentiles for Written Composition Variables, Grade Three

Score	Impr.	Cont.	Devl.	Sent.	Vocab.	Conv.	Averg.
.7	1		1	1	1		
1.0	3		3	3	3	2	1
1.3	5	1	10	4	5	7	2
1.7	12	5	23	11	10	14	5
2.0	30	20	41	21	22	32	15
2.3	46	36	58	35	41	51	29
2.7	66	57	74	56	62	67	59
3.0	81	76	82	74	84	78	75
3.3	88	87	90	83	90	86	85
3.7	94	95	92	93	95	91	93
4.0	97	97	97	98	97	94	98
4.3	98	98	99	98	98	97	98
4.7	99	99		99	99	99	99

The inter-rater reliabilities for the four Grade Three teachers are presented in Table 32. Essentially, this is a statistical measure of the amount of agreement or disagreement among the raters who scored the papers. This type of study sheds light on the question of whether, in this type of scoring, there is adequate agreement among raters that the process can be considered reliable. These results indicate that, in terms of reliability, the raters did an excellent job. The data suggest that it would be safe to use two raters and, in the event of disagreement, involve a third head scorer who could cast a deciding vote.

Table 32

Reliability Study - Written Composition Marking, Grade Three

Grade	Scale	All Agree	All Within 1	Disagree
Three	General Imp.	32%	90%	10%
	Content	30%	88%	12%
	Development	21%	80%	20%
	Sentence	23%	83%	17%
	Vocabularly	32%	91%	9%
	Conventions	34%	96%	4%
Four and Five	General Imp.	32%	87%	13%
	Content	34%	89%	11%
	Development	24%	78%	22%
	Sentence	25%	80%	20%
	Vocabularly	39%	90%	10%
	Conventions	24%	81%	19%

Grades Four and Five. Tables 33 and 34 provide the overall district means and standard deviations for Grades Four and Five. Tables 35 and 36 provide the correlations among the various features examined, while Tables 37 and 38 show the percentile rankings for the various scores obtained by the students based on the holistic and analytic scoring. These Tables represent the combined scores of written compositions based on the narrative and persuasive modes, both of which should be familiar to students at these levels.

The results are much the same as those obtained for Grade Three. At the Grade Four level, a score of 3.0 (average of five components) would place the student in a percentile ranking from about 48-64 depending on the component under examination. At the Grade Five level, a score of 3.4 (average of five components) would place the student in a percentile ranking from about 37-55 depending on the component under examination.

Table 39 provides the inter-rater reliabilities for the nine Grade Four/Five raters. As with the Grade Three raters, these teachers did an excellent job, with no reliability score below .78.

Table 33

Overall District Results for Grade Four Written Composition

	Mean	Std. Dev.	Confidence Int.
General Impression	2.8	.8	2.6 to 3.2
Content	3.1	.9	3.0 to 3.3
Development	3.1	.9	3.0 to 3.3
Sentence Structure	2.9	1.0	2.8 to 3.0
Vocabularly	3.0	.7	2.9 to 3.2
Conventions	2.9	.9	2.8 to 3.0
Average of Five Analytic Components	3.0	.8	2.9 to 3.1

Table 34

Overall District Results For Grade Five, Written Composition

	Mean	Std. Dev.	Confidence Int.
General Impression	3.1	.9	3.0 to 3.2
Content	3.4	.9	3.3 to 3.5
Development	3.6	.9	3.5 to 3.7
Sentence Structure	3.3	.8	3.2 to 3.4
Vocabularly	3.4	.7	3.3 to 3.5
Conventions	3.3	.8	3.2 to 3.4
Average of Analytic Components	3.4	.7	3.3 to 3.5

Table 35

Correlations Among Written Composition Variables, Grade Four

	Impr.	Cont.	Devl.	Sent.	Vocab.	Conv.	Averg.
General Impression	1.00	.77	.76	.70	.74	.70	.85
Content		1.00	.66	.55	.76	.53	.80
Development			1.00	.79	.69	.76	.90
Sentence				1.00	.71	.83	.90
Vocabularly					1.00	.67	.87
Conventions						1.00	.88
Average of Analytic Components							1.00

Table 36

Correlations Among Written Composition Variables, Grade Five

	Impr.	Cont.	Devl.	Sent.	Vocab.	Conv.	Averg.
General Impression	1.00	.79	.72	.66	.74	.87	.84
Content		1.00	.66	.60	.83	.58	.86
Development			1.00	.72	.58	.73	.87
Sentence				1.00	.58	.79	.86
Vocabularly					1.00	.58	.82
Conventions						1.00	.86
Average of Analytic Components							1.00

Table 37

Percentiles for Written Composition Variables, Grade Four

Score	Impr.	Cont.	Devl.	Sent.	Vocab.	Conv.	Averg.
.7	1		1		1	1	
1.0	1	1	1	2	1	2	1
1.3	3	2	2	5	2	5	2
1.7	7	4	4	9	4	10	4
2.0	19	13	9	18	9	19	8
2.3	34	27	17	27	17	29	18
2.7	49	36	32	39	29	40	35
3.0	65	50	48	55	53	53	54
3.3	77	64	61	68	74	69	68
3.7	84	73	71	78	84	80	80
4.0	91	80	80	85	91	87	89
4.3	95	87	89	91	95	93	93
4.7	98	95	94	95	98	97	97
5.0	99	99	99	99	99	99	99

Table 38

Percentiles for Written Composition Variables, Grade Five

Score	Impr.	Cont.	Devl.	Sent.	Vocab.	Conv.	Averg.
1.3	2	1	1	1		1	
1.7	5	3	3	3		3	1
2.0	12	8	5	5	2	7	3
2.3	21	16	9	11	6	13	7
2.7	33	23	17	21	14	24	17
3.0	50	35	27	34	32	38	32
3.3	64	49	37	50	52	55	48
3.7	74	61	49	67	68	68	65
4.0	83	70	64	80	80	78	79
4.3	88	79	78	88	90	88	88
4.7	94	89	89	93	95	95	95
5.0	99	99	99	99	99	99	99

Table 39

Reliability Study - Written Composition Marking, Grades Four and Five

Grade	Scale	All Agree	All Within 1	Disagree
Three	General Imp.	32%	90%	10%
	Content	30%	88%	12%
	Development	21%	80%	20%
	Sentence	23%	83%	17%
	Vocabularly	32%	91%	9%
	Conventions	34%	96%	4%
Four	General Imp.	32%	87%	13%
	Content	34%	89%	11%
and	Development	24%	78%	22%
	Sentence	25%	80%	20%
Five	Vocabularly	39%	90%	10%
	Conventions	24%	81%	19%

Problem Two: Comparison of Holistic/Analytic Scoring, Grade Three, Discussion and Findings

Problem Two was concerned with comparisons of holistic and analytic scoring in terms of inter-rater reliability, cost effectiveness, and teacher satisfaction with these scoring techniques. The Action Plan and Timeline called for these comparisons to be completed by October, 1985, and these were followed well.

Inter-rater reliability data based on the June 1984 assessment of written composition at Grade Three are provided in Table 32. (See page 88.) Inter-rater reliability for holistic (general impression) scoring was .90, and inter-rater reliability ranged from .83 to .96 for the five features of the analytic scoring scale. Thus the reliability for both types of scoring was sufficiently high that both techniques can be considered reliable. There was no difference between the inter-rater reliability for holistic and analytic scoring that had any practical significance.

Cost effectiveness comparisons involved the collection of data by members of the Grade Three scoring team concerning the number of hours required to score the written compositions both holistically and analytically. These data reveal that the raters were able to score holistically about 40 compositions per hour, at an approximate rate of 1.5 minutes per paper. They were able to score analytically about 14 papers per hour, at an approximate rate of 4.3 minutes per paper.

Clearly, holistic scoring was more cost effective. However, it must be noted that analytic scoring requires examination of five features while holistic scoring requires the examination of only one. Analytic scoring provides considerably more information to teachers about students' writing skills than does holistic scoring. Such information may be essential in planning appropriate instructional programs in writing. Thus school divisions must decide if the information obtained from analytic scoring is sufficiently more useful than the information obtained from holistic scoring to warrant the cost differential.

Teacher perceptions of the relative values of holistic and analytic scoring are discussed in detail in the next section of this Report. The data concerning teacher perceptions were obtained from the responses (to a questionnaire and a structured interview) of Grade Three teachers who had been involved in both the 1983 and 1984 assessments. When asked if the information obtained from holistic scoring only was of value, 36% of the teachers felt that it was, 36% felt that it was not, and 27% provided no response. When asked if the information obtained from both holistic and analytic scoring was of value, 100% of the teachers agreed that it was. When asked which of the two techniques was of most value, 73% responded with analytic scoring and 27% with holistic. Thus it is clear that the majority of teachers prefer analytic to holistic scoring if a choice must be made, but all of the teachers feel both types of scoring are useful. (See Tables 44-46, pp. 109, 110.)

Problem Three: Comparison of Writing Modes, Grades Four/Five, Discussion and Findings

Problem Three asked if there were significant differences in achievement in written composition at Grades Four and Five depending on the mode of writing required. These modes were narrative and expository/persuasive. The Action Plan and Timeline called for this comparison to be completed in the Summer of 1984, which it was.

At the Grades Four/Five level, students were randomly assigned to write in either the narrative or the expository/persuasive mode. Two topics were provided, one on camping and one on becoming teacher for a day. Students were permitted to select the topic, but had no choice as to task.

Table 40 provides the results of an analysis of variance conducted by Dr. Tom Maguire. In all cases, the mean scores for Grade Five were significantly higher than those for Grade Four. Where the topic was found to be significant, the mean was higher for camping than for teacher for a day. Where the task was significant, the mean for narrative writing was higher than the mean for expository/persuasive writing. Thus there were significant differences in achievement in written composition based on writing mode.

Table 41 shows the significant interactions based on the analysis of variance. The topic/task (BC) interaction was significant in the cases of general impression, content, vocabulary, and average of analytic components. Generally, this resulted from an unexpectedly high value for the narrative compositions on the topic of teacher for a day.

Table 40

Analysis of Variance, Written Composition, Grades Four and Five
(Scores are listed out of 15 possible)

Variable	Source	df	Mean Square	F	Sig		
General Impression	Grade (A)	1	132.04	19.97	yes		Grade 4: 8.3
	Topic (B)	1	18.71	2.83			Grade 5: 9.2
	Task (C)	1	50.48	7.64	yes		Expository/Persuasive: 8.4
	AB	1	11.44	1.73			Narrative: 9.0
	AC	1	2.01	0.30			
	BC	1	44.94	6.80	yes	(see A below)	
	ABC	1	0.54	0.08			
	Within	547	6.61				
Content	Grade (A)	1	131.54	16.83	yes		Grade 4: 9.2
	Topic (B)	1	2.31	0.30			Grade 5: 10.2
	Task (C)	1	6.02	0.77			
	AB	1	2.24	0.29			
	AC	1	0.97	0.12			
	BC	1	147.40	18.86	yes	(see B below)	
	ABC	1	1.89	0.24			
	Within	547	7.81				
Develop	Grade (A)	1	239.44	33.87	yes		Grade 4: 9.4
	Topic (B)	1	79.06	11.19	yes		Grade 5: 10.8
	Task (C)	1	0.09	0.01			Teacher: 10.5
	AB	1	11.58	1.64			Camping: 9.8
	AC	1	19.03	2.69			
	BC	1	13.18	1.86			
	ABC	1	10.31	1.46			
	Within	547	7.07				
Sentence	Grade (A)	1	191.69	29.02	yes		Grade 4: 8.8
	Topic (B)	1	213.08	32.26	yes		Grade 5: 10.0
	Task (C)	1	274.96	41.63	yes		Teacher: 10.1
	AB	1	16.15	2.45			Camping: 8.8
	AC	1	4.28	0.65			Expository/Persuasive: 8.6
	BC	1	13.42	2.03			Narrative: 10.0
	ABC	1	8.92	1.35			
	Within	547	6.61				

Table 40 (cont'd)

Analysis of Variance, Written Composition, Grades Four and Five
(Scores are listed out of 15 possible)

Variable	Source	df	Mean Square	F	Sig	
Vocabularly	Grade (A)	1	197.72	43.82	yes	Grade 4: 8.9
	Topic (B)	1	27.76	6.38	yes	Grade 5: 10.1
	Task (C)	1	35.59	8.18	yes	Teacher: 9.8
	AB	1	7.32	1.69		Camping: 9.3
	AC	1	0.36	0.08		
	BC	1	58.77	13.50	yes (see C below)	
	ABC	1	0.77	0.16		Expository/Persuasive: 9.2
	Within	547	4.35			Narrative: 9.7
Conventions	Grade (A)	1	173.65	25.04	yes	Grade 4: 8.7
	Topic (B)	1	99.92	14.41	yes	Grade 5: 9.8
	Task (C)	1	17.48	2.52		Teacher: 9.7
	AB	1	24.52	3.54		Camping: 8.8
	AC	1	9.34	1.35		
	BC	1	9.29	1.34		
	ABC	1	0.88	0.13		
	Within	547	6.93			
Average of Analytic Components	Grade (A)	1	183.66	37.60	yes	Grade 4: 9.0
	Topic (B)	1	55.56	11.37	yes	Grade 5: 10.2
	Task (C)	1	347.54	7.11	yes	Teacher: 9.9
	AB	1	10.41	2.13		Camping: 9.3
	AC	1	2.58	0.53		
	BC	1	35.40	7.25	yes (see D below)	
	ABC	1	0.44	0.09		Expository/Persuasive: 9.3
	Within	547	4.89			Narrative: 9.8

Table 41

Significant Interactions, Grades Four and Five

	A Impression		B Content		C Vocabularly		D Average of Analytical Components	
	Exp.	Nar.	Exp.	Nar.	Exp.	Nar.	Exp.	Nar.
Teacher	8.2	9.5	8.7	10.2	9.0	10.3	9.3	10.4
Camping	8.5	8.6	10.0	9.4	9.3	9.3	9.2	9.3

RESEARCH CLUSTER THREE--TEACHER PERCEPTIONS

The following problems concerning the area of teacher perceptions were identified for investigation in the study:

1. What is the effect, as reported by teachers, of involvement in a product evaluation, on teaching emphases and behavior?
2. What is the perception of teachers regarding the relative values of a product-oriented evaluation and a process-oriented evaluation?
3. What is the perception of teachers regarding the relative values of holistic and analytic scoring of written composition?

These problems were based on three of the study's major purposes. These were:

1. an examination of teachers' perceptions concerning the effect of their involvement in a product-oriented evaluation of the language arts program;
2. an examination of teachers' perceptions regarding the relative value of product oriented and process-oriented program evaluation;
3. an examination of the relative value of two techniques for the scoring of written compositions.

The following Action Plan and Timeline for these problems were included in the Proposal.

Problems One and Two

(1) May 85	Design a questionnaire to be administered to all elementary teachers who have been involved in this evaluation and previous process-oriented program evaluations.
(2) June 85	Administer the questionnaire and analyze the data.

Problem Three

(1) May 84	Design a questionnaire to be administered to all Grade Three teachers relative to the perceived benefits of information received from holistic scoring as compared to analytic scoring.
(2) May 84	Design a structured interview format to be administered to all Grade 3 teachers involved in both holistic and analytic scoring relative to the perceived benefits.
(3) July 84	Administer instruments alluded to in (1) and (2) above and analyze data.

Problems One and Two: Process/Product Evaluations, Discussion

As noted earlier, Grande Prairie School District #2357 has had a tradition of process-oriented program evaluation, defined as that type of evaluation which examines inputs and processes rather than student achievement products and which is conducted by a team whose members are identified as having special knowledge in the area under examination. It is to be expected that teachers who have been involved in such evaluations would have developed perceptions as to their benefits or lack thereof.

The evaluation activities associated with this study have been product-oriented--that is, they have been concerned with the assessing of student outcomes. Research Cluster Three dealt with the perceptions of teachers who had been involved with both types of evaluation regarding their relative value. With the recent emphasis by Alberta Education on achievement testing, Diploma evaluations, and other types of product-oriented evaluations, an examination of these perceptions seemed appropriate.

Problem One was concerned with teachers' perceptions of the effect on teaching emphases and behavior of their involvement in a product evaluation. Problem Two was concerned with the relative values of the two types of evaluation. To deal with these problems, a questionnaire was developed and administered to 52 teachers at the elementary level who had been involved with product evaluations in this study and with process evaluations at other times within the school district. (See Questionnaire A.)

Two items on Questionnaire A were designed to provide answers to the question posed as Problem One. These items were: (1) What do you you see as the benefits of the Elementary Language Arts Product Assessment? and (2) What problems/concerns do you see with a product program evaluation of this nature? One item was designed to elicit answers to the question posed as Problem Two. This was: (3) In your opinion which of the two forms of program evaluation (the process-oriented program evaluation or the product-oriented program evaluation) is of most benefit (a) to you personally and (b) to the growth and improvement of school district programs? Demographic data in terms of teaching experience in Grande Prairie and involvement in the product evaluation were also collected through this questionnaire.

Problem One: Process/Product Evaluation, Findings

Table 42 provides the responses to the first two items on Questionnaire A. These items asked the teachers what they perceived as the benefits of and problems/concerns with a product-oriented program assessment.

A considerable number of responses indicated that involvement in a product-oriented program evaluation had a positive effect, especially in determining student strengths and weaknesses. In addition, by establishing benchmarks teachers were able to gain useful information for program planning. On the other hand, teachers were concerned that the product evaluation had the potential to constitute teacher evaluation and reflect negatively on them especially in light of individual student circumstances. The benefits, however, appeared to outweigh the concerns, although attention must be given to these concerns if product evaluations are to continue in this district.

Table 42

Product Assessment: Benefits and Problems/Concerns

Benefits	**Frequency**
Provides student, school, and teacher benchmarks	18
Indicates implications for program delivery	18
Improves assessment, diagnosis, and evaluation	16
Helps identify student strengths, weaknesses	13
Provides inservice opportunities	3
Validates teacher evaluations	2
Helps students learn to write tests	2
No benefits	3
No responses	3

Problems/Concerns	**Frequency**
Product Measures do not reflect individual circumstances	12
No problems	9
Promotes teaching to the test	8
Promotes over-emphasis on comparisons	8
Feedback for teachers and program is minimal	7
Results might be used to assess teaching competence	6
Increases stress on teachers and students	6
Does not evaluate teaching process	5
Time lag in receiving results	4
Time consuming	4
Product measures inappropriate	3
No response	0

Problem Two: Process/Product Evaluation, Findings

Table 43 provides the responses to the third item on Questionnaire A. This item asked the teachers to compare the process-oriented and product-oriented program evaluation techniques in terms of benefits to themselves and to the school district.

Concerning personal benefits, about 60% of the teachers were split in their perceptions of relative value, with half of that group preferring the product evaluation and the other half preferring the process evaluation. It should be noted here that 23 of the 52 teachers had participated as members of either the Product Steering Committee, Listening Test Development Committee, or Written Test Composition Marking Committee. However, no breakdown was made concerning which teachers supported which evaluation type. It is interesting to note that about 30% of the teachers preferred both types of evaluations, although that response was not specified on Questionnaire A.

Concerning district benefits, while 25% of the teachers felt that product evaluation would be of benefit to the district and 30% felt that process evaluation would be of benefit to the district, 33% preferred both types. More teachers felt that product evaluation would be of benefit to them personally than would be of benefit to the school district. This suggests that the school district should consider the use of both types of evaluation, and that further research might be done to determine the most appropriate circumstances for the application of each type of program evaluation.

Table 43

Personal Benefits of Product/Process Evaluations

Type	**Frequency**	**Percentage**
Product evaluation	16	31
Process evaluation	16	31
Both	15	29
No response	5	9

District Benefits of Product/Process Evaluations

Type	**Frequency**	**Percentage**
Product evaluation	13	25
Process evaluation	16	31
Both	17	33
No response	6	11

Problem Three: Holistic/Analytic Comparison, Discussion

Problem Three was concerned with teachers' perceptions of the relative values of holistic and analytic scoring of written composition. In order to deal with this problem, a questionnaire was designed which included three items: (1) Was the information gained from the 1983 marking of the written compositions holistically for general impression only of any benefit to you? If so, how? (2) Was the information gained from the 1984 marking of the written compositions holistically for general impression as well as for content, development, sentence structure, vocabulary, and conventions of any benefit to you? If so, how? and (3) In your opinion, which of the two marking processes was of most benefit to you? Why? Demographic data in terms of teaching experience in Grande Prairie and membership in the Grade Three written composition marking team were also elicited. (See Questionnaire B.) As well, a structured interview was carried out with four of these teachers.

Questionnaire B was administered to 11 teachers, all of whom had been involved in both holistic and analytic marking. Of these, four had been trained in these scoring techniques as members of the marking teams, while the others had learned about the scoring processes from the Teacher Marking Packages developed by the district. The four trained markers were involved in the structured interview.

Problem Three: Holistic/Analytic Comparison, Findings

Tables 44 to 46 present the teachers' responses to Questionnaire B. As noted earlier, the written composition assessment done in 1983 included holistic (general impression) scoring only, while the 1984 assessment included both holistic and analytic marking. As well, by late 1984, the Teacher Marking Package for Grade Three had been developed and distributed to teachers.

Of the 11 teachers who responded to Questionnaire B, about one third felt that the information obtained from holistic marking alone was useful, about one third thought it was not useful, and about one third provided no response. A further examination of the data provided in Table 44 indicated that the four teachers who felt the information was beneficial were the trained markers.

Of these 11 teachers, all felt that the information obtained from both holistic and analytic marking was useful, especially in terms of the identification of student strengths and weaknesses.

Concerning the relative value of holistic and analytic scoring, most of the teachers (73%) felt that analytic scoring was most beneficial. This is consistent with their perceptions as reported in Table 45. Clearly they felt that they gained more information from analytic scoring, and that they could use this technique more appropriately for instruction.

The Structured Interview was carried out with the four members of the Grade Three marking team who had been involved with both holistic and analytic scoring during the 1983 and 1984 assessments of written composition. The Interview included such questions as (1) Has your experience(s) marking holistically last year for general impression only and analytically this year for sub-component skills as well given you something you can apply in the classroom? (2) What, if anything, have you learned about the process of writing in these marking experiences; and (3) Given the premise that one of the purposes of this part of the project was to establish writing standards in our school/district, how do you feel each scoring technique has impacted on that goal?

These teachers' comments during the Structured Interview provided reinforcement for the opinions provided in Tables 42-46. Among their comments were these:

- It (the marking experience) makes our curriculum objectives much more clear.

- It helps in terms of evaluation for report card purposes.

- Once you have gone through 250 papers, you have a better idea of what students can do.

- It is surprising how many commonalities come through.

- Using analytic scoring will have more classroom application than just using holistic scoring because it will reflect back on where your teaching is strong and where it is weak and will say something about what you're actually doing in the classroom.

- A lot of times when I was marking I had the urge to hand the paper back to the kid and say to him, "This is how I think you can make your paper better."

- I think by including descriptive writing we are more aware that narrative isn't the only way to write.

- You can be more specific in your comments about remedial work, and show students specific areas they can work on.

- When you come to parent-teacher interviews you can say, "Your son or daughter's writing is excellent in these four areas but weak in this one." You can really pinpoint where the problems are.

- For report card marking I would like to use general impression scoring but for teaching I would like to look at the specific areas of writing.

- When I signed up to work (on the marking team) I wanted it to be helpful to me. I found it very helpful because I have always had trouble marking and knowing exactly what to look for.

- I think it broadens your perspective on writing and on the marking of writing. I always had the tendency in the past to focus on conventions and sentence structure but it (the marking) has forced me now to think of the kids' ideas and how they have gotten them arranged.

> - It has definitely been a valuable educational experience and once in a while it was even entertaining.

The opinions expressed by these teachers match those expressed in the literature by other teachers who have been involved with the various types of holistic scoring. It would appear that, when teachers are encouraged to take ownership in their own learning, they learn better and more quickly. If they perceive that the knowledge they are gaining is valid and relevant in their classrooms, so much the better. From a staff development perspective, involving teachers in their own learning about the evaluation of writing should add to their sense of self-efficacy and worth. This study has shown that indeed it did, and that once in a while it was even entertaining.

Table 44

Holistic/General Impression Marking Benefits (1983 Marking of Written Compositions)

	Yes #	Yes %	No #	No %	NR #	NR %
Was the information beneficial?	4	36	4	36	3	27

If so, how?	Frequency
Comparative purposes relative to other students in the district	1
Gave overall impression of student writing abilities	1
Established standards for marking, general impression	1
Resulted in greater emphasis on written work in the classroom	1

Table 45

Holistic/General Impression and Analytic Marking Benefits (1984 Marking of Written Compositions)

	Yes #	Yes %	No #	No %	NR #	NR %
Was the information beneficial?	11	100	-	-	-	-

If so, how?	Frequency
Helped to identify strengths and weaknesses of the program	5
Helped to identify strengths and weaknesses of the students	5
Established standards/consistency regarding what we should be looking for from both teachers and students	4

Table 46

Comparable Benefits, Holistic/Analytic Marking

	Holistic		Analytic	
	#	%	#	%
Which process was of most benefit?	3	27	8	73

Why?	Frequency
Additional information gained	3
Comparative and diagnostic benefits	1

QUESTIONNAIRE A

GRANDE PRAIRIE SCHOOL DISTRICT #2357

ELEMENTARY LANGUAGE ARTS PRODUCT ASSESSMENT PROJECT
TEACHER QUESTIONNAIRE

Over the past few years, teachers in the Grande Prairie School District have been involved in several process-oriented program evaluations. The Elementary Science, Art and Social Studies evaluations are the most recent examples of such evaluations. In this type of evaluation, the emphasis is on the factors relevant to program delivery. In addition, over the past three years teachers in the Grande Prairie School District have also been involved in an Elementary Language Arts product-oriented program evaluation. The focus in this type of evaluation is on student achievement. In this instance achievement in the writing, reading, spelling and listening program components in Language Arts was monitored.

One of the research questions to be investigated in this Language Arts Product Assessment project is:

> "What is the perception of teachers regarding the relative values of a product-oriented evaluation as opposed to a process-oriented evaluation?"

Your responses/comments to the questions below would be much appreciated.

1. What do you see as the benefits of the Elementary Language Arts Product Assessment?

   ____

   ____

   ____

2. What problems/concerns do you see with a product-oriented program evaluation of this nature?

   ____

   ____

   ____

3. In your opinion which of the two forms of program evaluations (the process-oriented program evaluation or the product-oriented program evaluation) is of most benefit:

   a) to you personally?

   ____

   ____

   ____

   b) to the growth and improvement of school district programs?

   ____

   ____

   ____

4. Did you teach in the Grande Prairie School District:

   a) in the 1982/83 school term? Yes____ No____

   If so, at what grade level? ____________

   b) in the 1983/84 school term? Yes____ No____

   If so, at what grade level? ____________

   c) What grade level are you teaching this school term?

   ____

5. Were you involved in any of the projects listed below?

   a) Listening Test Development Committee. Yes____ No____

   b) Written Composition Test Marking Teams. Yes____ No____

   c) Language Arts Product Evaluation Steering Committee. Yes____ No____

Please return your completed questionnaire by May 31 to:

Lorne Radbourne
c/o Central Office

Thank you.

QUESTIONNAIRE B

GRANDE PRAIRIE SCHOOL DISTRICT #2357

ELEMENTARY LANGUAGE ARTS PRODUCT ASSESSMENT PROJECT
GRADE THREE WRITTEN COMPOSITION TEST TEACHER QUESTIONNAIRE

In June 1983 and again in June 1984, the Grande Prairie School District did an assessment of the writing competence of its grade three students. The grade three students were asked to write in the descriptive mode. More specifically, each grade three class was required to write a description of a large poster sized color picture. The picture displayed nine hot air balloons in various stages of inflation with a background showing Bear Creek, trees, spectators and houses.

In 1983 the written compositions were scored holistically for general impression only. Holistic scoring is a process by which a written composition is assigned a mark in terms of how well it meets a predetermined set of criteria.

In 1984 each written composition was again scored holistically for general impression; but in addition, the papers were re-scored holistically in each of the five writing component areas. (content, development, sentence structure, vocabulary, conventions)

Both in 1983 and 1984 teacher written composition marking packages were prepared in order to share scoring criteria, classroom application suggestions and test result summaries with all grade three teachers in the Grande Prairie School District.

One of the research questions to be investigated in the Language Arts Product Assessment project we have been conducting in our District the past three years is:

> "What is the perception of teachers regarding the relative values of scoring written compositions holistically for general impression only and scoring written compositions holistically for general impression as well as content, development, sentence structure, vocabulary and conventions?"

Your responses/comments to the questions below would be much appreciated.

1. a) Was the information gained from the 1983 marking of the written compositions holistically for general impression only of any benefit to you?

   ____

   b) If so, how?

   ____

   ____

   ____

2. a) Was the information gained from the 1984 marking of the written compositions holistically for general impression as well as content, development, sentence structure, vocabulary and conventions of any benefit to you?

   ____

   b) If so, how?

   ____

   ____

   ____

3. a) In your opinion, which of the two marking processes was of most benefit to you?

   ____

   b) Why?

   ____

   ____

   ____

4. Were you a member of the grade three written composition marking team?

   Yes_____ No_____

5. Please circle the school term(s) that you have taught grade three in the Grande Prairie School District.

   1982/83

   1983/84

   1984/85

Please return your completed questionnaire by May 31 to:

Lorne Radbourne
c/o Central Office

Thank you.

RESEARCH CLUSTER FOUR--PARENT PERCEPTIONS

The following problem concerning parent perceptions of student achievement was presented for investigation in the study.

> What is the relationship between parental perceptions of student achievement in selected areas of language arts, and the actual achievement as determined by the product measures?

The major purpose on which this problem was based was: an examination of the relationship between parents' perceptions of student achievement in selected language arts areas and actual student achievement in these areas.

This Statement of Proposed Action appeared in the Proposal: The intended course of action relative to the above question would be to design and administer parent questionnaires at each grade level. Such questionnaires would list selected questions (or in the case of written composition, descriptors and samples of composition grade ratings) and would ask parents their perceptions of the appropriateness of the question for the grade level, and their estimate of the percentage of students who might be able to respond to the question. In addition, they would be asked questions of a general nature relative to student achievement. Results would be compared to actual achievement.

Instead of a questionnaire method, it was decided to use a direct action procedure to answer the question. Two language arts areas were selected for investigation, that of writing and reading. Two activities were undertaken in this regard.

Activity One: Written Composition, Discussion and Findings

In the first activity, parents, two from each of the district's six elementary schools, were invited to participate in a half-day marking session of Grade Four/Five written compositions. First, the parents were given a two hour training session to familiarize them with the procedures involved in the holistic (general impression) marking of written compositions and with the criteria used for the analytic marking of compositions: content, vocabulary, development, conventions, and sentence structure. Following this training session, each parent independently marked 4 of the 12 sample compositions selected for the activity. Thus, each paper was scored by three separate parent markers. Table 47 shows the holistic/general impression and analytic scores (on a 5 point scale) awarded by the three parent markers for each of the criteria for each composition. (It should be noted that the scores reported in both Tables 47 and 48 are the average marks provided by the parents and teachers.)

Table 48 shows the holistic/general impression and analytic scores given to each of the same papers by teams of three teachers using the same scoring criteria. A comparison of Tables 47 and 48 reveals that the scores given by the parents were generally consistent with those of the teachers. This is not particularly surprising in light of the fact that holistic scoring has been shown to be a reliable method of marking written compositions. Inter-rater reliability studies have found a high degree of agreement among raters who used holistic marking techniques. As holistic scoring is a process by which a written composition is assigned a mark in terms of how well it meets a pre-determined set of criteria, the marking process naturally becomes more objective.

Table 47

Summary of Parent Marker Scoring

Paper #	Holistic General Impression Scores	Analytic Scores					Average Mark (For Each) Paper
		Conventions	Sentence Structure	Vocabularly	Content	Development	
327	4.0	3.3	3.7	4.7	4.3	4.0	4.0
528	2.0	2.5	3.5	2.5	1.5	3.0	2.5
433	2.0	1.0	1.3	2.7	2.0	1.3	1.7
2	5.0	4.5	4.5	4.5	4.0	5.0	4.6
541	3.0	3.7	3.7	3.3	2.7	4.3	3.5
523	1.0	1.0	0.7	1.0	0.7	1.0	0.9
229	4.7	4.3	5.0	4.3	4.0	4.3	4.4
368	3.0	2.0	2.7	3.0	2.0	3.7	2.7
207	1.5	0	1.0	0	1.0	0.7	0.7
51	3.0	2.7	3.7	4.0	3.7	3.3	3.4
57	3.5	3.0	3.0	3.0	4.0	4.0	3.4
266	0.3	1.0	1.3	3.0	2.0	2.0	1.6
Average Score	2.8	2.4	2.8	3.0	2.7	3.0	2.8

*Note: The scores given under each marking component are the average marks awarded by three parent markers.

Table 48

Summary of Teacher Marker Scoring

Paper #	Holistic General Impression Scores	Analytic Scores					Average Mark (For Each) Paper
		Conventions	Sentence Structure	Vocabularly	Content	Development	
327	4.0	4.3	4.3	3.3	4.0	5.0	4.2
528	2.0	2.0	3.0	3.0	2.0	3.0	2.5
433	1.0	1.0	1.0	1.7	1.3	1.0	1.2
2	4.7	4.7	5.0	5.0	5.0	5.0	4.9
541	3.0	4.0	3.3	2.7	2.0	4.0	3.2
523	1.0	0.3	1.0	1.0	1.0	1.7	1.0
229	3.0	3.7	4.0	3.3	4.0	4.3	3.7
368	2.0	3.0	2.7	3.0	2.7	3.0	2.7
207	1.3	1.3	1.3	1.3	1.3	1.7	1.4
51	4.0	4.0	5.0	4.7	4.7	5.0	4.6
57	2.7	2.0	3.0	3.0	3.7	3.0	2.9
266	1.3	1.7	1.7	3.0	2.7	1.3	2.0
Average Score	2.5	2.7	2.9	2.9	2.9	3.2	2.9

*Note: The scores given under each marking component are the average marks awarded by three teacher markers.

Activity Two: Reading, Discussion and Findings

In the second activity, 12 parents, again two from each of the six elementary schools, were given copies of the Edmonton Public School Board's Grade Six Reading Test to review. Then they were asked to predict the percentage of Grade Six students in Grande Prairie School District #2357 who would score within the following three percentage ranges: (1) 80-100%; (2) 50-79%; and (3) 0-49%. These predictions were then compared with the actual percentages of students who achieved scores on the test within each of these percentage ranges. The findings are provided in Table 49.

These findings reveal that actual student performance was considerably higher than the parents had predicted, and reinforce the notion that parents may perceive student ability to be lower than it is. While this was a very small sample of parents, more work of this nature may help parents develop a more reasonable understanding of students function in schools.

Thus, in terms of the relationship of parent perceptions of student achievement and actual student achievement as determined by product measures, a stronger relationship was found between perceptions of student achievement in written composition and actual student achievement than was found between perceptions of student achievement in reading and actual student achievement.

It was interesting to note the enthusiasm and support demonstrated by parents for these activities. All of the parents were positive because they felt the activities afforded them an excellent opportunity to learn about and gain some appreciation for what the schools are attempting to accomplish with their children.

Perhaps one reason why parents and interested citizens are concerned about the products of our education system is that they do not have a clear conception of what is happening in schools and why. Educators might well be advised to make concerted efforts in this area in order to promote a better understanding of the goals, objectives, methodologies, and successes of our present education system.

Table 49

Summary of Parent Predictions and Actual Student Achievement On Grade Six EPSB Reading Test

Achievement Score Ranges	Percentage of Student Scores Within Each Range As Predicted by Parents	Actual Percentage of Students Scoring Within Each Range
80%-100%	19%	32%
50%-79%	61%	56%
0%-49%	20%	12%

IV. RESEARCH SUMMARY

This study, Evaluation of Selected Components of an Elementary Language Arts Program in a Small Urban School Jurisdiction, was conducted by Grande Prairie School District #2357 under contract with Alberta Education. The time span of the study was two years, from 1983 to 1985. Final data analyses and the Final Report were completed in 1986.

The study had as its major purposes: (1) the development and validation of achievement tests in the area of listening; (2) an investigation of the relative merits of two techniques for the scoring of written compositions; (3) an examination of teachers' perceptions regarding the relative value of product-oriented and process-oriented program evaluation; (4) an examination of teachers' perceptions concerning the effect of their involvement in a product-oriented evaluation of the language arts program; and (5) an examination of the relationship between parents' perceptions of student achievement in selected language arts areas and actual student achievement in these areas as determined by product measures.

In order to meet the purposes of the study, four research clusters were specified, each with its own set of problems. In this section, the research questions and the findings pertaining to these questions are summarized.

Research Cluster One--Listening

Problem One. Can student achievement in the listening dimension of the Alberta Elementary Language Arts Curriculum be assessed validly and reliably at the Grades One to Four levels?

To answer this question, listening achievement tests, one at each of Grades One to Four, were developed and administered by teachers in Grande Prairie

School District #2357. Statistical analyses of the data obtained from the test administrations were conducted. Two versions were pilot tested. A third and final version, the Grande Prairie Listening Tests (GPLT) was administered at each of the Grades One to Four levels in June, 1985.

The answer to the question is a qualified yes. The Grades One and Two tests were found to be both valid and reliable. The Grades Three and Four tests were found to require additional investigation as problems were identified with either reliability, validity, or both.

Problem Two. Do the conceptual domains of listening skills which underlie and form the basis of the construction of the Edmonton Public Listening Tests have empirical validity?

To answer this question, Battery A of the Edmonton Public Listening Tests, a Canadianized version of the British Tests of Listening Comprehension, was administered at the Grades Five and Six level in June, 1983, and June, 1984. Statistical analyses of the data obtained from these administrations were conducted.

The answer to this question based on these data is no. There was a lack of strong relationship among the test items and little evidence that the items clustered in the alignment suggested by the subtests. Thus it would seem that the traits measured by the subtests are not independent of each other, and that the conceptual domains of listening skills which form the basis of the construction of these tests do not have empirical validity.

Problem Three. What other language arts skills correlate most highly with listening skills?

To answer this question, a number of language arts variables were correlated with the total listening scores of the Grande Prairie Listening Tests and the Edmonton Public Listening Tests. These variables differed from grade to grade. In general, correlations were highest with reading test scores and verbal ability scores, and were lowest with spelling test scores and holistic scores on written compositions.

Research Cluster Two--Written Composition

Problem One. Can the analytic scoring techniques that have been applied at secondary school levels be used reliably for the rating of written compositions at Grades Three, Four, and Five?

The question of reliability of written composition ratings was examined through the use of both holistic (general impression) and analytic scoring (in this case for five features or components: content, development, sentence structure, vocabulary, and conventions). A Grade Three writing assessment involving holistic scoring only was conducted in June, 1983, and another involving both holistic and analytic scoring was conducted in June, 1984. As well, a Grade Four and Five assessment involving both holistic and analytic scoring was conducted in June, 1984.

The answer to this question was yes at all grade levels. Inter-rater reliability statistics revealed a high degree of agreement among raters, such that both the analytic and holistic scoring techniques could be considered reliable at Grades Three, Four, and Five.

Problem Two. How does analytic scoring compare to holistic scoring in Grade Three for each of the following dimensions: (a) inter-rater reliability; (b) cost effectiveness; and (c) teacher satisfaction (both in terms of those teachers involved and those who receive the information from scoring)?

(a) Inter-rater reliability. No differences of any practical significance were found between the inter-rater reliabilities for holistic and analytic scoring.

(b) Cost effectiveness. Holistic scoring was found to be more cost effective. The use of holistic scoring resulted in the marking of about 40 compositions per hour while the use of analytic scoring resulted in the marking of about 14 compositions per hour. However, it must be noted that holistic scoring required the reading of each composition for only one factor, while analytic scoring required the reading of each composition for five factors.

(c) Teacher satisfaction. See Research Cluster Three--Teacher Satisfaction.

Problem Three. Are there significant differences in achievement in written composition in Grades Four and Five depending on the mode of writing required?

The Grade Four and Five writing assessment involved two writing tasks, one narrative and one expository/persuasive, which were randomly assigned to students by their teachers. In all cases the mean scores for Grade Five were significantly higher than for those of Grade Four. Where the task (mode of writing) was significant, the mean for narrative writing was higher than the mean for expository/persuasive writing.

The assessment also involved two writing topics, one on camping and the other on becoming teacher for a day. Students were permitted to choose their topic. It is interesting to note that, where the topic was found to be significant, the mean was higher for camping than for teacher for a day.

Research Cluster Three--Teacher Perceptions

Problem One. What is the effect, as reported by teachers, of involvement in a product evaluation, on teaching emphases and behavior?

To answer this question and the question posed in Problem Two below, a questionnaire was designed and administered to 52 elementary teachers who had been involved in both product-oriented and process-oriented program evaluations.

Two items on the questionnaire were designed to elicit responses to Problem One: (1) What do you see as the benefits of the Elementary Language Arts Product Evaluation? and (2) What problems/concerns do you see with a product program evaluation of this nature?

Four benefits were mentioned 13 times or more by teachers: (1) provides student, school, and teacher benchmarks; (2) indicates implications for program delivery; (3) improves assessment, diagnosis, and evaluation; and (4) helps identify student strengths and weaknesses. Three other benefits mentioned by teachers received frequency counts of 3 or less. A longer list of problems/concerns was provided by teachers, with relatively lower frequency counts. Of these, four were mentioned seven times or more: (1) product measures do not reflect individual circumstances; (2) promotes teaching to the test; (3) promotes over-emphasis on comparisons; and (4) feedback for teachers and program is minimal. It is interesting to note that the response of "No benefits" had a frequency count of three, while the response of "No problems" had a frequency count of nine.

Problem Two. What is the perception of teachers regarding the relative values of a product-oriented evaluation and a process-oriented evaluation?

One item on the questionnaire mentioned above was designed to elicit responses to Problem Two: In your opinion which of the two forms of program evaluation (product-oriented or process-oriented) is of most benefit (a) to you personally and (b) to the growth and improvement of school district programs?

In terms of both personal and district benefits, about one third of the teachers preferred product evaluations, one third preferred process evaluations, and one quarter to one third preferred both types of evaluations. Slightly more teachers felt that product evaluation would be of more benefit to them personally than it would to the district as a whole.

Problem Three. What is the perception of teachers regarding the relative values of holistic and analytic scoring of written compositions?

To answer this question, a questionnaire was designed and administered to 11 Grade Three teachers who had been involved in both types of scoring. The questionnaire consisted of three items: (1) Was the information gained from the 1983 marking of the written compositions holistically (only) of any benefit to you? If so, how? (2) Was the information gained from the 1984 marking of the written compositions holistically for general impression as well as for content, development, sentence structure, vocabulary, and conventions (analytically) of any benefit to you? If so, how? and (3) In your opinion, which of the two marking processes were of most benefit to you? Why? Of the 11 teachers, four had been members of the marking team that rated the written compositions during both the 1983 and 1984 assessments, while the other seven had received information about the scoring processes and the results of the assessments.

Concerning the 1983 assessment, about one third of the teachers felt that information obtained from holistic scoring was beneficial, about one third felt the information was not beneficial, and about a third had no response. It is interesting to note that the third who responded positively consisted wholly of those four teachers who had served on the marking team.

Concerning the 1984 assessment, all of the teachers felt that the information obtained from the combination of holistic and analytic scoring was beneficial.

Concerning the relative value of holistic and analytic marking, about three quarters of the teachers felt that analytic scoring was of most benefit, while about one quarter felt that holistic scoring was of most benefit. Those who preferred the analytic scoring technique indicated that they did so because they felt it had greater classroom application since it provided information about five aspects of writing rather than just a general impression of pieces of writing.

A Structured Interview was also held with the four Grade Three teachers who had served as members of the marking team for both assessments. The opinions expressed by the teachers reinforced the findings of the questionnaire survey.

RESEARCH CLUSTER FOUR--PARENT PERCEPTIONS

Problem: What is the relationship between parental perceptions of student achievement in selected areas of language arts, and the actual achievement as determined by product measures?

Two activities were undertaken to answer this question, one concerned with writing and the other with reading.

In Activity One, a group of parents and teachers trained in holistic and analytic scoring provided these two types of scores for a common set of Grade Four/Five written compositions. The parent and teacher scores for both types of marking were found to be quite consistent, with the average scores being the closest.

In Activity Two, a group of parents was asked to predict the percentages of students whom they thought would score in three percentile ranges on the Grade Six Edmonton Public School Board Reading Test. These predicted percentages were then compared with the actual percentages of students who achieved scores in these percentile ranges on the test. The actual student percentages within these ranges were found to be considerably higher than the parents' predictions, revealing a weaker relationship between parent perceptions of student achievement in reading than in written composition.

V. CONCLUSIONS AND RECOMMENDATIONS

RESEARCH CLUSTER ONE--LISTENING

Conclusion 1.1: The Grande Prairie Listening Tests at the Grades One and Two levels appear to be valid and reliable measures of the objectives on which they were based while the Tests at levels Three and Four appear to have some problems with validity and reliability. It should be noted, however, that the tests involved stimulus selections based on different topics, formats, and language types, each of which may have influenced the degree of student attention to the selections, questions, and answer options.

Recommendation 1.1.1: The Grande Prairie Listening Tests should be administered to additional samples of the school population at each grade level, and further analyses should be performed on the results with a view toward establishing acceptable performance criteria as well as reliability and validity at all grade levels.

Recommendation 1.1.2: The Grande Prairie Listening Tests should be lengthened and expanded in terms of items per objective and stimulus selections in order to determine the effects of attention and ability to deal with specific language types, formats, and stimulus selections.

Recommendation 1.1.3: Continuing efforts should be made to devise test items which will reveal developmental trends in the area of listening. This may involve the development of additional and perhaps more specifically stated listening objectives.

Conclusion 1.2: The Grande Prairie Listening Tests appear to be appropriate for use as initial tools for identifying students who have difficulty in listening.

Recommendation 1.2.1: When used as a screening device, the tests should be readministered to students who seem at risk on the first administration.

Recommendation 1.2.2: The tests should not be used as the sole device for identifying students' strengths and weaknesses in listening. Teachers' perceptions of students' listening ability as well as students' listening performance in real classroom contexts should be considered as well in identifying students' needs in this area.

Conclusion 1.3: Results of the administration of the Grande Prairie Listening Tests have relevance to the planning and implementation as well as the assessment of the instructional program in listening.

Recommendation 1.3.1: Test results should not be used for program evaluation at Grades Three and Four until additional research at these levels indicates appropriate reliability and validity.

Recommendation 1.3.2: Teachers at the various grade levels should develop appropriate listening instructional sequences, materials, and activities based on the objectives, language types, formats, and stimulus selections used in the Grande Prairie Listening Tests.

Conclusion 1.4: Professional development is enhanced when teachers are involved in the building of evaluation instruments within the context of a working group.

Recommendation 1.4.1: Teachers should be encouraged to work together to develop evaluation techniques and instruments in subject areas other than listening as well as to expand their work with the development of listening tests.

Conclusion 1.5: The use of the Edmonton Public Listening Tests at the Grades Five and Six levels does not appear appropriate based on data collected with the Grande Prairie population sample.

Recommendation 1.5.1: The Edmonton Public Listening Tests should be administered to larger samples of the school population at each grade level, and further analyses should be performed on the results with a view to determining the extent to which the findings of this study are replicated.

Recommendation 1.5.2: If the use of the Edmonton Public Listening Tests is not deemed appropriate after further investigation, teachers in the Grande Prairie School District #2357 should develop listening instruments similar to the Grande Prairie Listening Tests for use in Grades Five and Six.

RESEARCH CLUSTER TWO--WRITTEN COMPOSITION

Conclusion 2.1: Both holistic and analytic scoring techniques are appropriate for use at Grades Three, Four, and Five.

Recommendation 2.1.1: The appropriateness of using holistic and analytic scoring techniques at Grades One, Two, and Six should be investigated.

Recommendation 2.1.2: Continuing inservice in the use of these techniques should be held for teachers new to the district.

Conclusion 2.2: Holistic scoring is more cost effective than analytic scoring. On the other hand, analytic scoring provides more information to teachers, and when asked to choose, is preferred by them to holistic scoring.

Recommendation 2.2.1: District teachers and administrators should determine which type of scoring is most appropriate to the various types of instructional activities included in the writing curriculum.

Conclusion 2.3: Both writing tasks (modes) and topics influence student achievement in written composition.

Recommendation 2.3.1: A variety of writing modes, including narrative, descriptive, expository, and persuasive, should be included in the writing program at the various grade levels. As well, teachers should develop instructional activities in writing based on a variety of topics both more and less familiar to students at the elementary level.

RESEARCH CLUSTER THREE--TEACHER PERCEPTIONS

Conclusion 3.1: Teachers perceive product-oriented evaluations as having both benefits and problems. Benefits include the provision of information for improving evaluation and diagnosis as well as for improving program delivery. Problems include concerns that product measures may not reflect individual student circumstances and that product evaluations may lead to teaching to a test and/or to over-emphasis on both student and teacher achievement comparisons.

Recommendation 3.1.1: Research should be conducted on the extent to which the concerns voiced by teachers in this study reflect those of teachers at other grade levels in the district. If these concerns are common, administrators and teachers should work together to develop guidelines for the use of product evaluation feedback which will alleviate concerns.

Conclusion 3.2: Teachers perceive both product-oriented and process-oriented program evaluations as being beneficial to them personally as well as beneficial to the school district.

Recommendation 3.2.1: Teachers and administrators in Grande Prairie School District #2357 should determine the type of evaluation most appropriate to the various instructional programs with which the district is involved.

RESEARCH CLUSTER FOUR--PARENT PERCEPTIONS

Conclusion 4.1: Parent perceptions of student achievement in the language arts may or may not match teachers' perceptions or actual achievement as identified by product measures.

Recommendation 4.1.1: Further investigation should be undertaken of parent perceptions of student achievement in order to determine where contrasts with teacher perceptions or actual student achievement are greatest.

Recommendation 4.1.2: District teachers and administrators should develop and implement techniques and programs which promote parental understanding about the goals, objectives, methodologies, and successes of the school district.

REFERENCES

Crowhurst, M. (1977) The Effect of Audience and Mode of Discourse on the Syntactic Complexity of the Writing of Sixth and Tenth Graders. Unpublished Doctoral Dissertation, University of Minnesota.

Mullis, I. (1980) Using the Primary Trait System for Evaluating Writing. Denver, Colorado: National Assessment of Educational Progress.

Nyberg, V. and Nyberg, A. (1982) Alberta Essay Scales: Models. Edmonton: University of Alberta.

Odell, C. and Cooper, L. (1977) Evaluating Writing. Urbana, Illinois: National Council of Teachers of English.

Plattor, E. (1977) Assessing Listening Skills--An Examination of Selected Research Reports. Unpublished Report to the Minister's Advisory Committee on Student Achievement (MACOSA). Edmonton: Alberta Education.

Plattor, E., Unruh, W., Muir, L. and Loose, K. (1979) Test Development for Assessing Achievement in Listening and Speaking. Edmonton: Alberta Education.

Prater, D. and Padia, W. (May, 1983) "Effects of Modes of Discourse on Writing Performance in Grades Four and Six," Research in the Teaching of English, Volume 17, Number 2.

Shuman, R. (1977) Questions English Teachers Ask. New York: Rochelle Park.

Smulyan, L. (1984) Collaborative Action Research: Historical Trends. Unpublished Paper, Annual Meeting, American Educational Research Association.

Swanson, C. (1986) Teachers as Listeners: An Exploration. Unpublished Paper, Annual Meeting of the International Listening Association.

Temple, C. and Gillet, J. (1984) Language Arts: Learning Processes and Teaching Practices. Toronto: Little, Brown and Company, 1984.

Wagner, K. (1983) The Effect of Abstraction Level on Syntactic Structures in Written Composition at the Sixth Grade Level. Unpublished Masters Thesis, University of Victoria.

Wagner, K. (1983) Proposal: Evaluation of the Elementary Language Arts Program in Small School Jurisdictions Through Product Assessment. Grande Prairie: Grande Prairie School District #2357.

GRANDE PRAIRIE SCHOOL DISTRICT 2357

LANGUAGE ARTS PRODUCT EVALUATION

THE EVALUATION OF WRITTEN COMPOSITION

GRADE THREE

TEACHER MARKING PACKAGE

November 1984

Scorers: Ray Buziak, Hillside Community School
Jane McGarty, Alexander Forbes School
Marg Ridgeway, Hillside Community School
Jean Zatelny, Parkside Elementary School

Consultant: Tom Maguire, University of Alberta
Supervisors: Keith Wagner, Deputy Superintendent
Lorne Radbourne, Acting Deputy Superintendent

The contents of this Teacher Marking Package are not intended to imply, either explicitly or eitherwise, Alberta Education policy regarding the evaluation of students' writing skills or the evaluation of teaching performance.

ACKNOWLEDGEMENTS

The District would like to thank all the people who contributed to this project: the teachers who administered the writing task, the scorers, the consultant and supervisors and the clerical staff. The improvement of written composition is a continuing goal of this District, and this project has not only provided some baseline data to help determine whether we are achieving that goal; it has also contributed to the professional growth of all involved.

D. R. Taylor
Superintendent of Schools

TABLE OF CONTENTS

PAGE

I. Background Information 2

II. Scoring Criteria and Examples

A: The Criteria 6
B: Applying the Criteria 12

III. Classroom Application 14

IV. Results Summary 15

V. Appendices

A: Answer Guide 17
B: Scoring Descriptors for Gr. 3 Narrative 18
C: Scoring Descriptors for Gr. 4 Persuasive/Expository . 23

I. BACKGROUND INFORMATION

In June 1984 the Grande Prairie School District did an assessment of the writing competence of its grade three students. The grade three students were asked to write in the descriptive mode. More specifically, each grade three class was required to write a description of a large poster sized color picture. (50cm by 75cm). The picture showed nine hot air balloons in various stages of inflation. The background showed Bear Creek , trees, spectators and houses. The identical poster was distributed to each school along with the following specific directions to teachers:

1. Background

This test is part of a total language arts assessment research project in the District. The purposes of the test are as follows:

A. To develop a statement outlining the criteria/standards by which descriptive writing produced by Grade 3 students can be evaluated.

B. To determine the number of students in the District who write to these standards as determined by a committee of teachers.

C. To determine the reliability of holistic committee scoring procedures at this level.

D. To compare holistic scoring which produces only one general impression score with that which produces scores in component areas of writing skills such as content, development, sentence structure, vocabulary and conventions on such bases as:

i. Time and cost
ii. Teacher satisfaction
iii. Reliability

District results from this test will be shared with everyone in the District. School results will be shared only with the school to which they pertain.

Attempts will be made to publish representative compositions receiving excellent, good, fair and poor grades and descriptions of them. These should prove useful to all teachers in setting realistic expectations for their students in the future.

The specific District objective to which this relates is as follows:

"Given a wide range of sensory experiences with an object, the student writes a description of a maximum of 100 words using appropriate vocabulary to identify a minimum of 6 attributes."

2. Administration Time Permitted

It is intended that adequate time be provided so that no student is penalized because of time pressure. Therefore, even though very few students are likely to use the full time allotment, they should be provided a minimum of two consecutive class periods, preferably at a time in which they are fresh (For example, 9:00-10:20 a.m. or 1:00 - 2:20 p.m.).

Alternate quiet activities such as recreational reading or worksheet assignments should be provided for those who finish early.

3. Use of Posters

You have been provided two or three color posters. Place them in your classroom so that all the students can easily see one of the posters from where they are seated. You might wish to divide the students into groups for this purpose. Try to avoid crowding of desks to prevent copying or "borrowing" of ideas.

Students may have questions about the posters. "What are those?" etc. Do not answer any such questions. Simply explain that they are to "paint a word picture of everything they can" in the scene. Students may also want to leave their desks for a "close look". This should not be permitted either, as there should be adequate detail to allow for their descriptive abilities without providing this assistance.

4. Procedure

A. Group your students and posters as indicated in III above, so that all can easily see a poster.

B. Make sure that all students have pencils and an eraser and a clear desk.

C. Hand out the student test sheets, and read through the instructions with the students.

D. The I.D. Number area on the tests is not for the students' use, or for yours. Tell the students to ignore it.

E. If students ask the usual "how long should it be?" questions, tell them that it can be as long as they want - there are no limits. However, it has to say enough to "paint a complete word picture" of the scene.

F. Students can't use dictionaries or other aids. You should not supply them with words.

G. Students may use manuscript or cursive writing, whichever they are most comfortable with.

H. When students are clear about the instructions, have them begin. Circulate around the class to see that they have begun, but do not provide assistance.

After approximately 30 minutes, allow students a "stretch break" of one to five minutes.

I. As students complete their work, collect it from them and make sure they have recorded their names properly.

J. Group the completed tests alphabetically and forward them to Keith Wagner at Central Office.

The following specific instructions were given to students:

THIS IS A TEST TO SEE HOW WELL YOU CAN DESCRIBE A PICTURE IN WRITING. DO YOUR VERY BEST WORK.

Look at the posters that your teacher will show you. This is a picture of a scene in Bear Creek Park here in Grande Prairie.

Imagine that you are on a hill above the scene. Write about the scene so that someone who is not there can form a picture of it in his mind. Don't tell a story. Write a description. Paint a picture with your words. Use your senses. Tell about what you can see. Tell about the colors and shape and sizes. Imagine what you might hear and smell, and tell about that too.

You can make your description as long as you want. Whoever reads what you write should get a good picture of the scene. Use enough words and sentences to tell everything you can about the scene.

Before you begin, here are some rules:

1. The description must be your own. You can't talk to anyone else.

2. The words that you use must be your own. You can't ask your teacher for words or spellings. Don't worry too much about your spelling. Just do the best you can.

3. You should use a pencil for your writing. If you want to change anything, erase it neatly and then write it the way you wanted it to be.

The tests were then collected, assigned a random identification number and scored by a team of grade three teachers during the first week of July. The papers were first scored holistically for a general impression score. Approximately forty papers per hour were scored in this manner.

Holistic scoring is a process by which a written composition is assigned a mark in terms of how well it meets a predetermined set of criteria. In order to complete this task the team of teachers established the criteria for a descriptive passage based on a scale of 0-5 with 0 being the lowest score awarded and 5 being the highest score given. The specific criteria for this is included in the next section of this document. Each member of the team read each paper and assigned it a mark which was recorded on a separate marking sheet. The mark was not put on the paper. Thus markers did not know the identity of the author or the mark awarded by other markers.

These papers were then rescored again in a holistic manner but this time in each of five writing components areas. Criteria were established for these five sub-components:

Content
Development
Sentence Structure
Vocabulary
Conventions

The criteria for each of these components was based on the descriptors established for the June 84 Grade 6 provincial language arts exam. They can be found in the next section of this package. The same marking procedure as for general impression was followed. Using this approach, an average of 14 papers per hour were scored completely for all five sub-components.

II. SCORING CRITERIA AND EXAMPLES

1. The Criteria

GENERAL IMPRESSION SCORING DESCRIPTORS FOR DESCRIPTIVE WRITING MODE

5 Excellent

- exceptionally developed detail
- precise vocabulary
  - vocabulary displays a variety of new and interesting words
  - comparisons create vivid impressions
- displays exceptional thought and organization
- shows some evidence of style
- a dominant impression may be evident
- few mechanical errors
  - punctuation
  - sentence structure
  - capitalization
  - spelling

4 Very Good

- well developed detail
- some precise vocabulary
- displays good evidence of thought and organization
- some mechanical errors - but these don´t interfere with readability or meaning
- also includes Number 5 papers with many mechanical errors.

3 Average

- displays some thought and organization
- sufficient detail
- length is adequate to complete the task
- uses vocabulary appropriate to grade level
- mechanical errors interfere somewhat with the message and the readability.

2 Poor

- length is inadequate to complete the task
- lack of detail
- vague vocabulary - dull uninteresting words
- lack of thought and organization
- overall impression of disorder due to jumbled arrangement of ideas
- irrelevant details
- mechanical errors greatly interfere with message

1 Little or No Communication

- mechanical errors interfere with meaning to the extent that the composition is nearly illegible
- completely off topic

0 Insufficient

- too little writing exists for judgement to be made

CONTENT SCORING DESCRIPTORS
FOR DESCRIPTIVE WRITING MODE

5 Exceptional

- specific details used to describe setting and activities
- creates an atmosphere through the use of the senses
- creates a vivid overall impression and gives clear physical descriptions
- good use of imagery
- captures a dominant impression or sense of style

4 Proficient

- some specific details to describe setting
- some general sense of atmosphere is created
- appeals to most of the senses
- use of imagery is evident

3 Satisfactory

- evidence of specific appropriate details
- some attempt to create an atmosphere or overall impression
- attempts to use imagery

2 Limited

- few appropriate details
- very little attempt to create atmosphere
- limited appeal to senses
- very little use of imagery

1 Poor

- no appropriate details
- setting is not developed
- no sense of atmosphere
- no imagery

0 Insufficient

- too little writing exists for judgement to be made

DEVELOPMENT SCORING DESCRIPTORS
FOR DESCRIPTIVE WRITING MODE

5 Exceptional

- displays coherent thought and organization
- there is evidence of paragraphing
- organized sequence of descriptions and events
- shows excellent sense of beginning and closure

4 Proficient

- displays good evidence of thought and organization
- good sense of beginning and closure
- may have some slight confusion in flow of ideas

3 Satisfactory

- descriptions are in generally coherent sequence
- some sense of closure is evident
- some disorganization of ideas

2 Limited

- limited sense of sequencing the descriptions
- absence of sense of closure
- weak sense of organization

1 Poor

- no sequencing
- no closure
- no evidence of organization

0 Insufficient

- too little writing exists for judgement to be made

SENTENCE STRUCTURES SCORING DESCRIPTORS FOR DESCRIPTIVE WRITING MODE

5 Exceptional

- good variety of sentence structures, type, length is used
- controlled used of co-ordination
- sentence fragments if evident are used for effect

4 Proficient

- some variety in sentence structure, type and length
- little over-use of co-ordination
- few sentence fragments

3 Satisfactory

- little variety in sentence structure, type and length
- some over-use of co-ordination
- some sentence fragments evident

2 Limited

- most sentences are simple sentences
- little variety in length and structure
- definite use of co-ordination
- may have many sentence fragments

1 Poor

- sentences are immature and repetitious
- almost exclusive use of co-ordination
- sentence fragments impede meaning

0 Insufficient

- too little writing exists for a judgement to be made

VOCABULARY SCORING DESCRIPTORS FOR DESCRIPTIVE WRITING MODE

5 Exceptional

- specific, concrete, interesting words have been selected to create vivid images and precise details
- denotative meanings are accurate and effective

4 Proficient

- frequent use of specific concrete words adds clarity to the detail created
- denotative meanings are most frequently accurate and effective

3 Satisfactory

- some use of specific, concrete words
- some use of general words
- denotations are mostly correct

2 Limited

- few specific concrete words
- most words are general
- some inaccuracy of meaning

1 Poor

- only vague, general words are used
- restricted choice of words
- inaccuracy of meaning

0 Insufficient

- too little writing exists for judgement to be made

CONVENTIONS SCORING DESCRIPTORS
FOR DESCRIPTIVE WRITING MODE

5 Exceptional

- the communicative power of the composition is enhanced because of careful form, spelling, usage, punctuation and capitalization and neatness of writing.

4 Proficient

- communication is clear because of essentially correct form, spelling, usage, punctuation and capitalization
- few errors in proportion to length

3 Satisfactory

- some errors in form, spelling, punctuation, usage and capitalization but communication is adequate

2 Limited

- frequent errors in spelling, punctuation, usage and capitalization reduce communication
- work is not neatly done

1 Poor

- very weak in communication due to incorrect spelling, no punctuation and capitalization and poor grammar
- poor printing/writing make it very hard to read

0 Insufficient

- too little writing exists for a judgement to be made

2. Applying the Criteria

Included here are ten grade three papers on which to practice the practical application of this scoring method. Read each paper and by referring back to the descriptors assess a mark from 0-5 for each of the areas. You must remember to check the criteria for each sub-component before reading the paper to mark for that component. As a means of comparison, you can find the scores that the marking team awarded each paper by looking in the appendix section of this document.

It is important to refer always to the descriptors and mark according to how the paper meets those standards. Try not to compare papers to ones read previously.

Also, when marking for sentence structure you may have to mentally put in the correct punctuation - if the sentence structure is there the paper will be penalized in the conventions section for lack of punctuation, but shouldn't be penalized twice.

MARKING SHEET

Paper #	General Impression	Content	Develop.	Sentence Structure	Vocabulary	Conventions
77						
29						
36						
109						
233						
125						
113						
134						
176						
18						

I.D. NUMBER 77

In this picture half of Bear Creek is surrounded by eight balloons. On the left side of the picture is a fully inflatted balloon: It has yellow, orange, red, pink, blue and green stripes. The top and bottom of this balloon is black. The balloon right behind this balloon is partly inflatted. It has yellow, green and orange stripes. The more farther away balloons are all inflatted except for one. All the balloons are striped. The cars, trucks and vans are all attachec to the balloons. On the green grass hunded and hundreds of people are waiting for the race to begin.

I.D. NUMBER 77

Even Bear, Creek is waiting. It's running south with the wind. Right now the workmen are pumping air into the two uninflatted. There. Now for the count down, 5, 4, 3, 2, 1 GO!

Seven ballons

I.D. NUMBER 29

Seven bruitsful ballons at a park. They are big and colerful. One has black on the top and black on the botom and in the mitle of the ballon has lots of difrent colers. The other ones are lots of difrent colers. They have a big basket on the botom of the ballon and the ballon is made from materel and the basket is made from bamboow. Lot and lots of poeple were looking at the ballons and there are cars and trucks. There is water in front of the ballons and there is trees, broshes and grass. The ballons are on the grass. You can berley see the houses and you can berley see the trees. The balloons are as size as a house and the people are as snall as a

turn over.

figer nail, And it is a suny butfull day. The end

I.D. NUMBER 36

it is a good of the
balomns it is a nies one.
nieser whan the ather
Pating's I wiuth I
was a teater. of theis
cassroom and of theis.
cass too But I'm
Sill a Child But
I wiuth I was the,
bigist in my famly,
I wiuth. I was 19 year
old So I can be a
teater a teater is fun
I bat I wiuth that,
I was in a art bolomn
too with my famly,
and whan I was
little bady I wanti
in one and that is,
ol I want in a
art ballomns So SinSt
thin I do not go in one

I.D. NUMBER 36

maybe nxst. Summer ,
I was go. in a art bolomn.
all by my siv are with.
my sisrer and borer are.
may be my mom and dad and ,
may be with .my ante ,

thea .

I.D. NUMBER 109

Colored Raindrops

Here are lots of hot air balloons and alot of busy people. The balloons all together look like rainbow colored raindrops falling from the sky. Six of the brightly colored. hot air balloons are ready to take off. Three of them are half blown up. The positioning of the balloons are two in the lower left hand corner, seven at the top of the picture coming downward in a line. Two balloons that aren't blown up are out of line. They sound like a big giant huffing and puffing. The balloons stand beside a baby blue creek. The creek dividesBear creek into two parts. The balloons are standing on one part and a whole bunch of people are standing on the other part. Beside the balloons are trucks with people in, helping them get the balloons going. The trucks look very small be—

turn over

109

side the large hot air balloons.

The color of the grass is a heathy color of green. On the grass stand many many people. They all look like very colorful dots or sticks. In the middle of the grass stands a dark green clump of trees. On the edge of the clump is a little white building. It has about three windows, a chimney stack and a door. In the lower right hand corner is a very dark spot. The darkness is made from the shade of balloons. In the dark spot there are a few popular trees growing At the back there is a nice big group of trees. Behind it there are houses and buildings. The sky looks like a a strip of very pale blue paper. This is a very happy scene.

I.D. NUMBER 233

theirs people theirs. balloons thars trees. theirs trucks a rive and bushes and houses a trell theirs rope holding the balloons theirs numbers on the balloons the air smell and gas and fire and smock ai 8 balloons and grass and closs and some rocks a brich a parcking lote a raid and water and weeds.

I.D. NUMBER 125

See. There's a creek looks like a thik ribbon. Not all of the ballons are up yet, some colers are, Black as the bigining of night, Yellows of clear days, Green as newly grass in the summer, Blue as 9:00 skys, Natral blue, Rosy pink, Red as blood, Orange as an Orange, Yellow again, And Black at the end, lots of people, veaclues, A brige, In the background is a city, lots of trees, It's morning, pine trees, oak trees, One of the ballons says; C-GYZ1, I see lots of posts one of the fallons has an orange on top, with red, yellow, And green colers on it, Theres a person on the frige, theres another frige for cars Sound's A long wosh lots of those woshing sounds, the creek is flowing it goes shhhhhhhhh the wind swaying in the trees sounds like rustuling back and fro. Some people are wharing sorts, And some people are wering pants.

I.D. NUMBER 125

fed The balloone feel like; a blown up materal, rubber, Soft, non furry, Non hard, delacit, feels warm, when up in air feels windy, very cofofting.

I.D. NUMBER 113

Not too long aga a group of people got together. and they decided they would all go to Bear Creek Park with their hot air balloons. When they got there they all set up their balloons. Wow! What a sight! There were brown ones black ones, red ones. Every colour you can think of! There were all different shapes and sizes. When they got back it was

I.D. NUMBER 113

about 6:00 pm. When it was over everybody went home. None of them could wait until they could do it again!

I.D. NUMBER 134

Balloon Park

There are lots of spruce trees and bushes. There's a stream to wade in. There is lots of shade from the trees. There are lots of pretty sents from the flowers. Some of the grass is green and other parts are yellow still. There some sand to play in. There are some houses in the background. Every balloon has its own special number or I.di on it. The balloons have lots of colors like black, yellow, green, blue, red, pink and purple. Some people are watching and two are riding and one is driving. Summer and Fall are the best times to go balloon riding. I think. There are cars park right close to the balloons. As soon as all of the balloons are ready the race starts.

I.D. NUMBER 176

There were events at the park. There were eight hot air balloons. I was on ahill above the scene. My favorate balloon was theone that was bright yellow like the sun, with pretty orange stripes going down the way. It made funny sounds with the fire thing that looked like a burner off our barbaque. Anyway the fire made the balloon go up and up so that I couldn't see it any more. But that was o.k. because I Just sellected another one. And than I went up and up tell my mother and Father couldn't see me any more. And Ithink my mom started to cry and my dad was yelling come back. It was quite hot in the basket but I could stand it when we got really nigh up the people looked like insects, and the trees looked like carot tops, and the other

I.D. NUMBER 176

balloons birds eggs. Pretty small ones to. I'm glad they made baskets on the balloons, or else I would be broiling because I would be at the top of the baloon by now. But by now I was afraid. I was going to fall in the river, that looked like a puddle from were I was standing.

I.D. NUMBER 18

The pitcher is: tooken in
bear creek. in this pitcher
ther are hot air ballooms in
the park 7 ballooms ready
to lanch off. Theres also some
ballms that geting ready to
go up. And theres lots of
green trees. The trucks that
brout the ballooms there
are waiting to take them back.
And lots of . shade And
people and if I was there I
would smell the nice warm
wind. And nice blue sky
and the blue streams and
Theres blake yellow blue and
pink orege and green brown and
purple and you could
see somme of Grande
Prairie thats if you
come on top of
the hill theres shady.

I.D. NUMBER 18

spots here and there
and men are holding
balloons down becuse
there going up some are
laing down.

here are the last
two words of this story
the end.

III. PRACTICAL CLASSROOM APPLICATIONS

The holistic approach to marking identified components of student writing has several practical classroom applications.

It can assist the teacher in diagnosing strengths and weaknesses both in student performance and program delivery. This scoring method provides the teacher with very useful and appropriate feedback for reporting student progress and in planning individual program strategies. Having a standard set of criteria leads to a more objective and consistent evaluation of student writing.

Teacher marking time can be more effectively utilized by providing a greater focus for the evaluation of student writing. It is not necessary to mark every writing assignment for all components. If you have, for example, taught a lesson on the use of specific versus general words, you could mark the written assignment using only the vocabulary criteria.

Another application of the criterion is for students use in evaluating their own papers. It would be possible to put the criteria in a more simplified form so that students could gain some appreciation for the elements of good writing. They could grade their own performance and see how they could improve their written composition based on the marking criteria. Although students would probably experience some problem initially, this approach could be employed to develop skills in proofreading and editing both a student´s own work or the work of others.

IV. RESULTS SUMMARY

1. Overall District Results - Grade Three Written Composition

	Mean	Std. Dev.	Confidence Int.
General Impression	2.6	.7	2.5 - 2.7
Content	2.8	.7	2.7 - 2.8
Vocabulary	2.6	.7	2.6 - 2.7
Conventions	2.6	.8	2.5 - 2.7
Development	2.4	.8	2.3 - 2.5
Sentence Structure	2.7	.8	2.6 - 2.8
Average of 5 Components	2.6	.6	2.6 - 2.7

2. Correlation Among Written Composition Variables

	Impr.	Cont.	Devl.	Sent.	Vocab.	Conv.	Averg.
Impression	1.00	.72	.60	.68	.61	.68	.77
Content		1.00	.66	.64	.76	.62	.86
Development			1.00	.64	.65	.55	.83
Sentence				1.00	.67	.71	.87
Vocabulary					1.00	.58	.85
Conventions						1.00	.83
Average of Sub-Comp							1.00

A correlation coefficient is an index of the linear relationship between two variables. A perfect correlation relationship is 1 (as one variable goes up the other goes up as well) but a correlation greater than .5 is considered significant.

3. Percentiles for Written Composition Variables

Score	Impr.	Cont.	Devl.	Sent.	Vocab.	Conv.	Averg.
.7	1		1	1	1		
1.0	3		3	3	3	2	1
1.3	5	1	10	4	5	7	2
1.7	12	5	23	11	10	14	5
2.0	30	20	41	21	22	32	15
2.3	46	36	58	35	41	51	29
2.7	66	57	74	56	62	67	59
3.0	81	76	82	74	84	78	75
3.3	88	87	90	83	90	86	85
3.7	94	95	92	93	95	91	93
4.0	97	97	97	98	97	94	98
4.3	98	98	99	98	98	97	98
4.7	99	99		99	99	99	99

4. Reliability Study - Written Composition Marking

Grade	Scale	All Agrees	All Within 1	Disagree
Three	General Imp.	32%	90%	10%
	Content	30%	88%	12%
	Development	21%	80%	20%
	Sentence	23%	83%	17%
	Vocabulary	32%	91%	9%
	Conventions	34%	96%	4%
Four and Five	General Imp.	32%	87%	13%
	Content	34%	89%	11%
	Development	24%	78%	22%
	Sentence	25%	80%	20%
	Vocabulary	39%	90%	10%
	Conventions	24%	81%	19%

Dr. Tom Maguire of the University of Alberta carried out the above inter-rater reliability study. Essentially, this is a statistical measure of the amount of agreement or disagreement among the raters who scored the papers. This type of study sheds light on the question of whether, in this type of scoring, there is adequate agreement among raters that the process can be considered reliable. Dr. Maguire´s conclusion was that in terms of reliability, the judges or raters had done an excellent job. He suggested that it would be safe to use two raters, and in the event of disagreement, a third head scorer could cast a deciding vote.

V. APPENDIX A - ANSWER GUIDE

Following are the marks that the marking team gave each paper.

Paper	General Impression	Content	Develop	Sentence Structure	Vocabulary	Conventions
77	4	3	3	4	4	4
29	3	3	3	3	2	3
36	1	1	1	1	1	1
109	4	5	5	4	4	5
233	2	2	1	1	1	1
125	3	3	2	2	4	2
113	2	2	3	4	2	4
134	3	3	2	3	3	4
176	3	4	2	3	3	3
18	2	3	2	2	2	1

V. APPENDIX B - SCORING DESCRIPTORS FOR NARRATIVE WRITING MODE

REPORTING CATEGORY: CONTENT

(Selecting Details Appropriate to Purpose)

SCORE		DESCRIPTION OF PERFORMANCE
5	EXCEPTIONAL	Events are plausible within a context that is clearly established by the writer. Events and actions are connected implicitly to character motivation. Many precise and appropriate details establish characters and events even though experiences may be of an everyday nature.
4	PROFICIENT	Most events are plausible within a context that is clearly established by the writer. Events and actions are sometimes connected to character motivation. Many appropriate details establish characters and events even though experiences may be of an everyday nature.
3	SATISFACTORY	Some events are plausible within a context that is clearly established by the writer. Events and actions are infrequently connected to character motivation. Some appropriate details establish characters and events even though experiences may be of an everyday nature.
2	LIMITED	Few events are plausible within a context that is vaguely established by the writer. Events and actions are rarely connected to character motivation. Few appropriate details establish characters and events.
1	POOR	Events may be plausible but a context is unclear. There is a lack of appropriate detail.
0	INSUFFICIENT	Too little writing exists for a judgement to be formed.

- Taken from Grade 3 Provincial Language Arts Test Specifications.
- That which the student chooses to write about. This includes the WHO, WHAT, WHERE, and WHEN of a story.
- Details selected by the student will be either descriptive or narrative and associated with characters or events.

REPORTING CATEGORY: DEVELOPMENT
(Organizing Details into a Coherent Whole)

SCORE		DESCRIPTION OF PERFORMANCE
5	EXCEPTIONAL	Events have been placed in a coherent and recognizable sequence. The story's unity is strengthened by details about character and actions. Digressive details, if present, do not interfere with the development of the story. Appropriate closure has been achieved.
4	PROFICIENT	Events have been placed in a coherent sequence. The story's unity is sometimes supported by details about characters and actions. Digressive details, if present, do not interfere with the development of the story. Closure has been achieved.
3	SATISFACTORY	Events have been placed in a generally coherent sequence. Digressive details begin to interfere with the story's development. Closure has been attempted.
2	LIMITED	A sequence of events can be detected, but coherence is not achieved. Digressive details interfere with the unity of the story. Closure, if attempted, is unsuccessful.
1	POOR	No coherent sequence of events is apparent. Digressive details, if present, interfere greatly with the unity of the story. A sense of closure is missing.
0	INSUFFICIENT	Too little writing exists for a judgement to be formed. Writing that has been awarded an "Ins" for CONTENT is insufficient.

The method by which the student chooses to organize the content. This includes the sequence of events by which the student organizes the story into a coherent whole. The sequence may involve ordering by cause and effect, but more usually involves a sequence ordered by time and/or by place.

REPORTING CATEGORY: SENTENCE STRUCTURE
(Structuring Sentences Appropriately)

SCORE		DESCRIPTION OF PERFORMANCE
5	EXCEPTIONAL	Appropriate and/or purposeful variation in sentence type, length, and structure is evident. Co-ordination and subordination have been used appropriately. Sentence fragments and/or run-on sentences, if present, do not impede meaning.
4	PROFICIENT	Some appropriate and/or purposeful variation in sentence type, length, and structure is evident. Co-ordination and subordination are used appropriately but co-ordination is predominant. Sentence fragments and/or run-on sentences, if present, do not impede meaning.
3	SATISFACTORY	Occasional appropriate and/or purposeful variation in sentence type, length, and structure is evident. Co-ordination is used extensively but some subordination is present. Sentence fragments and/or run-on sentences, if present, do not impede meaning.
2	LIMITED	Little appropriate and/or purposeful variation in sentence type, length, and structure is evident. Co-ordination has been overused, sometimes inappropriately. Sentence fragments and/or run-on sentences, if present, impede meaning.
1	POOR	Co-ordination has been used almost exclusively and inappropriately. Sentence fragments and/or run-on sentences, if present, impede meaning.
0	INSUFFICIENT	Too little writing exists for a judgement to be formed. Writing that has been awarded an "Ins" for CONTENT is insufficient.

The forms of the sentences that the student uses. The category, sentence structure, includes the types of sentences, co-ordination (i.e., linkage of clauses, e.g., "...and so..." or "...but..."), subordination (e.g., "...because..." or "...then..."), and the arrangement within a sentence (e.g., subject/verb/subject). Sentence fragments and/or run-on sentences are considered to be part of "sentence structure" rather than "conventions."

REPORTING CATEGORY: VOCABULARY
(Selecting and Using Words and Expressions Appropriately)

SCORE		DESCRIPTION OF PERFORMANCE
5	EXCEPTIONAL	Precise and specific verbs, nouns, and/or modifiers have been used appropriately to create clear images.
4	PROFICIENT	Some specific verbs, nouns, and/or modifiers have been used appropriately to create clear images.
3	SATISFACTORY	Few specific verbs, nouns, and/or modifiers have been used appropriately to create clear images, but general words are varied and correct.
2	LIMITED	General verbs, nouns, and/or modifiers have been used correctly. Images are vague.
1	POOR	General verbs, nouns, and/or modifiers have been used, often incorrectly or repetitively. Images are unclear.
0	INSUFFICIENT	Too little writing exists for a judgement to be formed. Writing that has been awarded an "Ins" for CONTENT is insufficient.

The words chosen by the student. The vocabulary category considers the precision and clarity of word choice (e.g., "mumbled" instead of "said," "canoe" instead of "boat", "wicked" instead of "bad").

REPORTING CATEGORY: CONVENTIONS
(Following the Conventions of Written Language Appropriately)

SCORE		DESCRIPTION OF PERFORMANCE
5	EXCEPTIONAL	Control of spelling, punctuation, and capitalization facilitates clear communication. Misspellings are easily decipherable. Dialogue, if present, may not be punctuated properly.
4	PROFICIENT	General control of spelling, punctuation, and capitalization facilitates clear communication. Misspellings are generally decipherable. Dialogue, if present may not be punctuated properly.
3	SATISFACTORY	Some control of spelling, punctuation, and capitalization facilitates communication. Misspellings are generally decipherable. Dialogue, if present, may not be punctuated properly.
2	LIMITED	Lack of control of spelling, punctuation, and capitalization generally interferes with communication. Misspellings are often decipherable. Dialogue, if present, may not be punctuated properly.
1	POOR	Lack of control of spelling, punctuation, and capitalization severely interferes with communication. Misspellings are generally undecipherable. Dialogue, if present, may not be punctuated properly.
0	INSUFFICIENT	Too little writing exists for a judgement to be formed. Writing that has been awarded an "Ins" for CONTENT is insufficient.

The way in which the student uses standard conventions of language. This includes the use of standard spelling, punctuation, and capitalization.

V. APPENDIX C

GENERAL IMPRESSION SCORING DESCRIPTORS
FOR GRADE 4/5 PERSUASIVE/EXPOSITORY WRITING MODE
(Can be adapted for your grade level)

5 Excellent

- exceptional clarity of communication
- has an evident, developed style
- creativity and specificity of detail which is suited to the purpose and relevant to the topic, including the stating of most reasons in the persuasive composition.
- vocabulary is specific, descriptive, vivid, and connotative
- exceptional thought and organization, including an evident beginning, middle, and ending
- very few errors of convention relative to the length

4 Very Good

- clarity evident, but atmosphere or style may not be found consistently throughout the paper
- appropriate amount of detail, including the use of many reasons in the persuasive composition some precise vocabulary
- displays good evidence of thought, and a beginning, middle, ending sequence
- some mechanical errors, but not so many that they interfere with readability or meaning
  (this category may also include #5 excellent compositions that have many mechanical errors)

3 Average

- communicates satisfactorily, length is adequate to complete task
- satisfactory detail suited to purpose, including an attempt made at giving details in a persuasive paragraph
- uses vocabulary appropriate to topic and purpose
- mechanical errors interfere somewhat with the message and the readability.

2 Weak

- length is inadequate to complete the task
- lack of sufficient detail, inclusion of some irrelevant details
- creates an overall impression of disorder, and lacks a clear ending
- vague vocabulary
- mechanical errors greatly interfere with meaning

1 Poor

- length is inadequate to complete task
- cannot tell what the purpose or task is
- very few details and several of them are irrelevant
- mechanical errors interfere with meaning to the extent that the composition is nearly illegible

0 - No real communication at all

CONTENT

5 Exceptional

- choices and reasons are plausible
- paper represents exceptional thought

4 Proficient

- choices are plausible, and reasons are given for most choices
- paper represents a good deal of thought

3 Satisfactory

- most choices are plausible, and some are supported with reasons
- paper reflects some thought

2 Limited

- some choices are plausible
- very few or no reasons given for choices
- paper represents little thought

1 Poor

- most choices are implausible
- no reasons given to support choices
- no real thought evident in the paper

0 - too little writing exists for judgement to be made.

DEVELOPMENT

5 Exceptional

- displays coherent thought and organization, with obvious evidence of categorization and/or superordination/subordination
- supported to some degree by paragraphing and/or by transitionals (for example, because of this, etc.)
- includes evidence of introduction and closure

4 Proficient

- displays coherent thought and organization with some evidence of categorization and/or superordination/subordination

3 Satisfactory

- displays little coherent thought and organization
- categorization barely evident

2 Limited

- categorization not evident
- poorly organized

1 Poor

- no evident organization at all
- rambling and hard to follow

0 - too little writing exists for judgement to be made.

SENTENCE STRUCTURE

5 Exceptional

- variety of sentence type, length, and structure is used for effects such as emphasis
- coordination has been controlled, and subordination is used appropriately
- sentence fragments, if present, are used for effect

4 Proficient

- variety is evident
- coordination is seldom overused
- subordination is often used appropriately
- there are few inadvertent fragments

3 Satisfactory

- some variety evident, but coordination may be overused
- subordination is successfully attempted
- fragments are in evidence, but do not impede meaning

2 Limited

- little variety and some awkward structures
- overdependence on coordination
- subordination, if used, is inappropriate
- fragments are frequent and impede meaning

1 Poor

- sentences are immature and there are many repetitious patterns
- coordination is used almost exclusively
- fragments are common and impede meaning

0 - too little writing exists for judgement to be made.

VOCABULARY

5 Exceptional

- occasional specific concrete words selected to create vivid images or precise details
- word meanings are accurate and effective

4 Proficient

- some use of specific or concrete words adds clarity to the detail created
- most word meanings are accurate and effective

3 Satisfactory

- some words have been selected appropriately but general or abstract words are often used where specific or concrete words would have been more effective
- some word meanings may be inaccurate or ineffective

2 Limited

- general words are usually used where some specific words would have been more effective
- many word meanings may be inaccurate or ineffective

1 Poor

- words convey only vague or general meanings

0 - too little writing exists for judgement to be made.

CONVENTIONS

5 Exceptional

- communicative power is enhanced because of careful spelling, grammar, punctuation, and capitalization

4 Proficient

- communication is clear because of essentially correct spelling, grammar, punctuation, and capitalization

3 Satisfactory

- communication is adequate because of generally correct spelling, grammar, punctuation, and capitalization

2 Limited

- communicative power is reduced because of incorrect spelling, grammar, punctuation, and capitalization

1 Poor

- communicative power is very weak because of errors in spelling, grammar, punctuation, and capitalization

0 - too little writing exists for judgement to be made.

GRANDE PRAIRIE SCHOOL DISTRICT #2357

LANGUAGE ARTS PRODUCT EVALUATION

GRADE THREE WRITTEN COMPOSITION

June 1984

TEACHER'S
MANUAL

I. Background

This test is part of a total language arts assessment research project in the District. The purposes of the test are as follows:

A. To determine the number of students in the District who write to standards determined by a committee of teachers.

B. To develop a descriptive statement of the characteristics of descriptive writing produced by Grade 3 students.

C. To determine the reliability of holistic committee scoring procedures at this level.

D. To compare holistic scoring which produces only one global score with that which produces scores in component areas of writing skills such as fluency, organization, sentence structure, and mechanics on such bases as:

1. Time and cost
2. Teacher satisfaction
3. Reliability

District results from this test will be shared with everyone in the District. School results will be shared only with the school to which they pertain.

Attempts will be made to publish representative compositions receiving excellent, good, fair and poor grades and descriptions of them. These should prove useful to all teachers in setting realistic expectations for their students in the future.

The specific District objective to which this relates is as follows:

"Given a wide range of sensory experiences with an object, the student writes a description of a maximum of 100 words using appropriate vocabulary to identify a minimum of 6 attributes."

II. Administration Time Permitted

It is intended that adequate time be provided so that no student is penalized because of time pressure. Therefore, even though very few students are likely to use the full time allotment, they should be provided a minimum of two consecutive class periods, preferably at a time in which they are fresh (For example, 9:00-10:20 a.m. or 1:00 - 2:20 p.m.).

Alternate quiet activities such as recreational reading or worksheet assignments should be provided for those who finish early.

III. Use of Posters

You have been provided two or three color posters. Place them in your classroom so that all the students can easily see one of the posters from where they are seated. You might wish to divide the students into groups for this purpose. Try to avoid crowding of desks to prevent copying or "borrowing" of ideas.

Students may have questions about the posters. "What are those?" etc. Do not answer any such questions. Simply explain that they are to "paint a word picture of everything they can" in the scene. Students may also want to leave their desks for a "close look". This should not be permitted either, as there should be adequate detail to allow for their descriptive abilities without providing this assistance.

IV. Procedure

1. Group your students and posters as indicated in III above, so that all can easily see a poster.

2. Make sure that all students have pencils and an eraser and a clear desk.

3. Hand out the student test sheets, and read through the instructions with the students.

4. The I.D. Number area on the tests is not for the students´ use, or for yours. Tell the students to ignore it.

5. If students ask the usual "how long should it be?" questions, tell them that it can be as long as they want - there are no limits. However, it has to say enough to "paint a complete word picture" of the scene.

6. Students can´t use dictionaries or other aids. You should not supply them with words.

7. Students may use manuscript or cursive writing, whichever they are most comfortable with.

8. When students are clear about the instructions, have them begin. Circulate around the class to see that they have begun, but do not provide assistance.

   After approximately 30 minutes, allow students a "stretch break" of one to five minutes.

9. As students complete their work, collect it from them and make sure they have recorded their names properly.

10. Group the completed tests alphabetically and forward them to Keith Wagner at Central Office.

PREAMBLE - In July 1984, four teachers worked as a team to evaluate approximately 250 Grade 3 written compositions. The compositions were a description of a hot air ballooning scene. The first task was to assess each composition for overall general impression on a scale of 0 to 5. To facilitate this task, the marking team used descriptors developed the previous year and outlined below. The team also developed descriptors for each of 5 sub-components: content, development, sentence structure, vocabulary, and conventions. All of the descriptors are outlined below:

GENERAL IMPRESSION SCORING DESCRIPTORS

5 Excellent

- exceptionally developed detail
- precise vocabulary
  - vocabulary displays a variety of new and interesting words
  - comparisons create vivid impressions
- displays exceptional thought and organization
- shows some evidence of style
- a dominant impression may be evident
- few mechanical errors
  - punctuation
  - sentence structure
  - capitalization
  - spelling

4 Very Good

- well developed detail
- some precise vocabulary
- displays good evidence of thought and organization
- some mechanical errors - but these don't interfere with readability or meaning
- also includes Number 5 papers with many mechanical errors.

3 Average

- displays some thought and organization
- sufficient detail
- length is adequate to complete the task
- uses vocabulary appropriate to grade level
- mechanical errors interfere somewhat with the message and the readability.

2 Poor

- length is inadequate to complete the task
- lack of detail
- vague vocabulary - dull uninteresting words
- lack of thought and organization
- overall impression of disorder due to jumbled arrangement of ideas
- irrelevant details
- mechanical errors greatly interfere with message

1 Little or No Communication

- totally illegible
- no response
- completely off topic

CONTENT SCORING DESCRIPTORS

5 Exceptional

- specific details used to describe setting and activities
- creates an atmosphere through the use of the senses
- creates a vivid overall impression and give clear physical descriptions
- good use of imagery
- captures a dominant impression or sense of style

4 Proficient

- some specific details to describe setting
- some general sense of atmosphere is created
- appeals to most of the senses
- use of imagery is evident

3 Satisfactory

- evidence of specific appropriate details
- some attempt to create an atmosphere or overall impression
- attempts to use imagery

2 Limited

- few appropriate details
- very little attempt to create atmosphere
- limited appeal to senses
- very little use of imagery

1 Poor

- no appropriate details
- setting is not developed
- no sense of atmosphere
- no imagery

0 Insufficient

- too little writting exists for judgement to be made

DEVELOPMENT SCORING DESCRIPTORS

5 Exceptional

- displays coherent thought and organization
- there is evidence of paragraphing
- organized sequence of descriptions and events
- shows excellent sense of beginning and closure

4 Proficient

- displays good evidence of thought and organization
- good sense of beginning and closure
- may have some slight confusion in flow of ideas

3 Satisfactory

- descriptions are in generally coherent sequence
- some sense of closure is evident
- some disorganization of ideas

2 Limited

- limited sense of sequencing the descriptions
- absence of sense of closure
- weak sense of organization

1 Poor

- no sequencing
- no closure
- no evidence of organization

0 Insufficient

- too little writing exists for judgement to be made

SENTENCE STRUCTURES SCORING DESCRIPTORS

5 Exceptional

- good variety of sentence structures, type, length is used
- controlled used of co-ordination
- sentence fragments if evident are used for effect

4 Proficient

- some variety in sentence structure, type and length
- little over-use of co-ordination
- few sentence fragments

3 Satisfactory

- little variety in sentence structure, type and length
- some over-use of co-ordination
- some sentence fragments evident

2 Limited

- most sentences are simple sentences
- little variety in length and structure
- definite use of co-ordination
- may have many sentence fragments

1 Poor

- sentences are immature and repetitious
- almost exclusive use of co-ordination
- sentence fragments impede meaning

0 Insufficient

- too little writing exists for a judgement to be made

VOCABULARY SCORING DESCRIPTORS

5 Exceptional

- specific, concrete, interesting words have been selected to create vivid images and precise details
- denotative meanings are accurate and effective

4 Proficient

- frequent use of specific concrete words adds clarity to the detail created
- denotative meanings are most frequently accurate and effective

3 Satisfactory

- some use of specific, concrete words
- some use of general words
- denotations are mostly correct

2 Limited

- few specific concrete words
- most words are general
- some inaccuracy of meaning

1 Poor

- only vague, general words are used
- restricted choice of words
- inaccuracy of meaning

0 Insufficient

- too little writing exists for judgement to be made

CONVENTIONS SCORING DESCRIPTORS

5 Exceptional

- the communicative power of the composition is enhanced because of careful form, spelling, usage, punctuation and capitalization and neatness of writing.

4 Proficient

- communication is clear because of essentially correct form, spelling, usage, punctuation and capitalization
- few errors in proportion to length

3 Satisfactory

- some errors in form, spelling, punctuation, usage and capitalization but communication is adequate

2 Limited

- frequent errors in spelling, punctuation, usage and capitalization reduce communication
- work is not neatly done

1 Poor

- very weak in communication due to incorrect spelling, no punctuation and capitalization and poor grammar
- poor printing/writing make it very hard to read

0 Insufficient

- too little writing exists for a judgement to be made

GRANDE PRAIRIE SCHOOL DISTRICT 2357

LANGUAGE ARTS PRODUCT EVALUATION

THE EVALUATION OF WRITTEN COMPOSITION

GRADE FOUR/FIVE

TEACHER MARKING PACKAGE

November 1984

Scorers: Merv Sylven, Hillside Community School
Ben Kijek, Alexander Forbes School
Don Safron, Alexander Forbes School
Angus Sutherland, Swanavon Elementary School
Larry Walton, Swanavon Elementary School
Marlyss Moller, Swanavon Elementary School
Jean Duplessis, Avondale Elementary School
George Crisfield, Crystal Park School
Sharon Odell, Harry Balfour School
Consultant: Tom Maguire, University of Alberta
Supervisors: Keith Wagner, Deputy Superintendent
Lorne Radbourne, Acting Deputy Superintendent

The contents of this Teacher Marking Package are not intended to imply, either explicitly or eitherwise, Alberta Education policy regarding the evaluation of students' writing skills or the evaluation of teaching performance.

ACKNOWLEDGEMENTS

The District would like to thank all the people who contributed to this project: the teachers who administered the writing task, the scorers, the consultant and supervisors and the clerical staff. The improvement of written composition is a continuing goal of this District, and this project has not only provided some baseline data to help determine whether we are achieving that goal; it has also contributed to the professional growth of all involved.

D. R. Taylor
Superintendent of Schools

TABLE OF CONTENTS

PAGE

I. Background Information 2

II. Scoring Criteria and Examples

A: The Criteria 8
B: Applying the Criteria 16

III. Classroom Application 18

IV. Results Summary 19

V. Appendices

A: Answer Guide 24
B: Scoring Descriptors for Gr. 3
Descriptive Writing Mode 25
C: Student Bloopers 31

I. BACKGROUND INFORMATION

In June, 1984 the Grande Prairie School District did an assessment of the writing competence of its grade four and five students. All grade four and five students were randomly assigned to write in either the narrative mode or the expository/persuasive mode. The following specific directions were give to teachers:

1. Background

This test of written composition is part of the Grande Prairie School District Language Arts Product Assessment Research Project. One of the research questions being investigated is whether students write as well in the expository persuasive mode as they do in the narrative mode. Consequently, students in your class have been randomly assigned to do either Writing Task A (expository/persuasive) or Writing Task B (narrative). Besides answering the above research question, other side benefits of this project include the following:

A. You will get some feedback regarding how well students in your class performed in the various skills that make up written composition, compared to the performance of a larger sample of students. This information may be valuable to you in making future decisions regarding your instructional emphasis.

B. Those teachers who participate in the scoring teams will gain experience in holistic methods of scoring. In addition, they will produce a holistic scoring handbook which might prove useful to you.

This test is meant to be administered over a two day period. This is to provide as realistic and desirable writing situation as possible. Students do their best writing when they have been provided some time to think about their ideas, and to discuss them with their peers.

2. Day One Instructions

In a double period block (approximately 70 to 90 minutes) hand out the test booklets and read through the two writing topics with the students. Explain that they may choose either topic, but regardless of the topic, they will be assigned by you to Writing Task A or Task B and will have no choice in that matter. Give them about 5 minutes to reread the topics, at the end of which they will be asked to select a topic. Then divide the class into groups of approximately 3 to 6 students who have chosen the same topic and have been assigned to the same Writing

Task. Tell the students that they will be given 15 minutes to discuss and share ideas about their Topic-Task. They are not to do any writing at this time. At the end of the 15 minutes, reorganize the class so that students work individually for the remainder of the period. (Approximately 40 minutes). Tell them that they can begin writing their compositions now, and will be given time to complete them and/or write a second draft the following day.

At the end of the period, collect their work to that point. They are not to be allowed to take their work home with them.

3. Day Two Instructions

On Day Two, students should be provided two consecutive class periods (from 70 to 80 minutes) in which to either complete their first drafts, edit, and revise, or write a second draft. Provide an alternative quiet activity for those who finish early.

The following guidelines should be shared with the students:

A. There is no required length, but marking consideration will be given to the extensiveness or completeness of ideas expressed.

The composition can be one or more paragraphs, depending on what they feel is necessary to express their ideas in an organized way.

B. Compositions should be handwritten, not printed.

C. Pens should be used - not pencils.

D. Students cannot use dictionaries or other aids, and the teacher (or friends) is not allowed to supply words, spellings, or other assistance.

4. Shipping Instructions

A. Make sure that students have filled in the required information on the front page of the test booklet, and that the booklets are stapled together securely with the pages in order. Students should not write their names anywhere on their compositions - only on the front page of the test booklet.

B. Bundle the test booklets up by class and forward them to Keith Wagner before June 30.

Thank you for your cooperation.

The following specific Instructions were give to students:

A. This is a test to see how well you can express your ideas in writing. Do your very best.

B. There are two topics. You will be given your choice of one of the two topics.

   Each topic has two possible writing tasks. Your teacher will assign you to either Task A or Task B.

C. You will be given a chance to discuss your topic with other students for a few minutes before you start writing. You will start your writing, let it set over night, and complete a second draft the next day.

D. Your work should be done:

   (1) In handwriting, not printing.
   (2) In pen, not pencil.

E. Once you start writing, you won´t be allowed to use a dictionary or other books, or to get help from your teacher or other students.

TOPIC I: CAMPING

Jason smiled broadly. It was the last hour of the last afternoon of the last day of school. Tomorrow he would be going camping with his mom and dad and sister. Jason thought of the fun they would have on this camping trip - a real camping trip with a tent and sleeping bags. They would even be cooking outside on a real fire.

Jason and his father were going to take their fishing poles along. Jason's mother and sister had been talking all week about the wild strawberries that grew in the open places in the woods near their camping spot. Jason's mouth watered as he thought of fresh trout cooked over an open fire and wild strawberries for dessert.

And there would be plenty for them all to do. They would be able to swim in the lake or ride horses from the stable down the road. Even walking through the woods watching for animals would be fun.

Jason thought of all of the possible adventures he could have exploring the woods with his sister. His smile increased. He could hardly wait for tomorrow to come.

Your teacher will assign you one of the tasks below. For your assigned task, write a composition that is well-organized and contains a variety of words, phrases, and sentences. Space is provided in this booklet for a first draft and a final copy.

TASK A

If you were planning a camping trip, what types of things would you want to take with you? Remember that you would need shelter, food, and clothing on you trip. Also, you might wish to include gadgets that would be useful for special purposes. Imagine that you would be camping for two days and could take only what you could put in the trunk of a car. Give reasons why you would take the things that you include.

TASK B

Write a story that tells what actually happens on Jason's camping trip, or on a camping trip that you went on. The story does not have to be true. Remember that an interesting story has a beginning, middle, and end. Also, a good story tells where the action took place, and who the story is about.

TOPIC II: TEACHER FOR THE DAY

All of the members of Mrs. Summer′s Grade 6 class were looking forward to the following day, but no one was quite as excited as Lori. Tomorrow was the day that Lori was to be "teacher for a day". Her classmates had elected her for this prestigious position, and Lori had spent most of the afternoon planning tomorrow′s activities with Mrs. Summer while the other children worked on their language arts assignment. Everything was ready now, and it looked like tomorrow would be a good day. The only thing that worried Lori was that Joel, the class trouble-maker and practical joker, had been giving her funny looks after class.

Your teacher will assign you one of the tasks below. For your assigned task, write a composition that is well-organized and contains a variety of words, phrases, and sentences. Space is provided in this booklet for a first draft and a final copy.

TASK A

If you were elected "teacher for a day", what kinds of activities would you plan for your classmates? Give good reasons for including the activities that you have chosen.

TASK B

Write a story that tells what happens on the day that Lori is "teacher for a day". Remember that a good story has a beginning, middle, and end, and tells who the story is about, and where the story takes place.

> The tests were then collected, assigned a random identification number and scored by a team of grade 4/5 teachers during the first week of July. The papers were first scored holistically for a general impression score.
>
> Holistic scoring is a process by which a written composition is assigned a mark in terms of how well it meets a predetermined set of criteria. In order to complete this task the team of teachers

established the criteria for both narrative and persuasive passages based on a scale of 0-5 with 0 being the lowest score awarded and 5 being the highest score given. The specific criteria for this is included in the next section of this document. Each member of the team read each paper and assigned it a mark which was recorded on a separate marking sheet. The mark was not put on the paper. Thus markers did not know the identity of the author or the mark awarded by other markers.

These papers were then rescored again in a holistic manner but this time in each of five writing components areas. Criteria were established for these five sub-components:

Content
Development
Sentence Structure
Vocabulary
Conventions

The criteria for each of these components was based on the descriptors established for the June 84 Grade 6 provincial language arts exam. They can be found in the next section of this package. The same marking procedure as for general impression was followed.

II. SCORING CRITERIA AND EXAMPLES

A. The Criteria

GENERAL IMPRESSION SCORING DESCRIPTORS FOR NARRATIVE AND PERSUASIVE WRITING MODES

5 Excellent

- exceptional clarity of communication
- has an evident, developed style
- creativity and specificity of detail which is suited to the purpose and relevant to the topic, including the stating of most reasons in the persuasive composition.
- vocabulary is specific, descriptive, vivid, and connotative
- exceptional thought and organization, including an evident beginning, middle, and ending
- very few errors of convention relative to the length

4 Very Good

- clarity evident, but atmosphere or style may not be found consistently throughout the paper
- appropriate amount of detail, including the use of many reasons in the persuasive composition some precise vocabulary
- displays good evidence of thought, and a beginning, middle, ending sequence
- some mechanical errors, but not so many that they interfere with readability or meaning
  (this category may also include #5 excellent compositions that have many mechanical errors)

3 Average

- communicates satisfactorily, length is adequate to complete task
- satisfactory detail suited to purpose, including an attempt made at giving details in a persuasive paragraph
- uses vocabulary appropriate to topic and purpose
- mechanical errors interfere somewhat with the message and the readability.

2 Weak

- length is inadequate to complete the task
- lack of sufficient detail, inclusion of some irrelevant details
- creates an overall impression of disorder, and lacks a clear ending
- vague vocabulary
- mechanical errors greatly interfere with meaning

1 Poor

- length is inadequate to complete task
- cannot tell what the purpose or task is
- very few details and several of them are irrelevant
- mechanical errors interfere with meaning to the extent that the composition is nearly illegible

0 - No real communication at all

CONVENTIONS - SCORING DESCRIPTORS FOR NARRATIVE AND PERSUASIVE WRITING MODES

5 Exceptional

- communicative power is enhanced because of careful spelling, grammar, punctuation, and capitalization

4 Proficient

- communication is clear because of essentially correct spelling, grammar, punctuation, and capitalization

3 Satisfactory

- communication is adequate because of generally correct spelling, grammar, punctuation, and capitalization

2 Limited

- communicative power is reduced because of incorrect spelling, grammar, punctuation, and capitalization

1 Poor

- communicative power is very weak because of errors in spelling, grammar, punctuation, and capitalization

0 - too little writing exists to make a judgement.

SENTENCE STRUCTURE - SCORING DESCRIPTORS FOR NARRATIVE AND PERSUASIVE WRITING MODES

5 Exceptional

- variety of sentence type, length, and structure is used for effects such as emphasis
- coordination has been controlled, and subordination is used appropriately
- sentence fragments, if present, are used for effect

4 Proficient

- variety is evident
- coordination is seldom overused
- subordination is often used appropriately
- there are few inadvertent fragments

3 Satisfactory

- some variety evident, but coordination may be overused
- subordination is successfully attempted
- fragments are in evidence, but do not impede meaning

2 Limited

- little variety and some awkward structures
- overdependence on coordination
- subordination, if used, is inappropriate
- fragments are frequent and impede meaning

1 Poor

- sentences are immature and there are many repetitious patterns
- coordination is used almost exclusively
- fragments are common and impede meaning

0 - too little writing exists to make a judgement.

VOCABULARY - SCORING DESCRIPTORS FOR NARRATIVE AND PERSUASIVE WRITING MODES

5 Exceptional

- occasional specific concrete words selected to create vivid images or precise details
- word meanings are accurate and effective

4 Proficient

- some use of specific or concrete words adds clarity to the detail created
- most word meanings are accurate and effective

3 Satisfactory

- some words have been selected appropriately but general or abstract words are often used where specific or concrete words would have been more effective
- some word meanings may be inaccurate or ineffective

2 Limited

- general words are usually used where some specific words would have been more effective
- many word meanings may be inaccurate or ineffective

1 Poor

- words convey only vague or general meanings

0 - too little writing exists to make a judgement.

CONTENT - SCORING DESCRIPTORS FOR
NARRATIVE WRITING MODE

5 Exceptional

- events are plausible and consistent with purpose
- specific details develop character, setting, atmosphere, or events

4 Proficient

- events are plausible
- appropriate details present a description of characters, setting, or events

3 Satisfactory

- most events are plausible
- several details used to describe characters, setting, or events

2 Limited

- many events are plausible
- a few details are used to describe characters, setting, or events

1 Poor

- events are implausible
- no details used to describe characters, setting, or events

0 - too little writing exists to make a judgement.

CONTENT - SCORING DESCRIPTORS FOR PERSUASIVE WRITING MODE

5 Exceptional

- choices and reasons are plausible
- paper represents exceptional thought

4 Proficient

- choices are plausible, and reasons are given for most choices
- paper represents a good deal of thought

3 Satisfactory

- most choices are plausible, and some are supported with reasons
- paper reflects some thought

2 Limited

- some choices are plausible
- very few or no reasons given for choices
- paper represents little thought

1 Poor

- most choices are implausible
- no reasons given to support choices
- no real thought evident in the paper

0 - too little writing exists to make a judgement.

DEVELOPMENT - SCORING DESCRIPTORS FOR NARRATIVE WRITING MODE

5 Exceptional

- displays coherent thought and organization
- supported to some degree by paragraphing and/or by transitionals (before, after this, meanwhile)
- contains organized sequence of description and events, including excellence of beginning, middle, and closure

4 Proficient

- displays coherent thought and organization
- shows distinct beginning, middle, and end

3 Satisfactory

- generally coherent and organized
- both beginning and closure are evident

2 Limited

- some lack of coherence and organization
- closure may not be evident

1 Poor

- no evident organization at all
- rambling and hard to follow

0 - too little writing exists to make a judgement.

DEVELOPMENT - SCORING DESCRIPTORS
PERSUASIVE WRITING MODE

5 Exceptional

- displays coherent thought and organization, with obvious evidence of categorization and/or superordination/subordination
- supported to some degree by paragraphing and/or by transitionals (for example, because of this, etc.)
- includes evidence of introduction and closure

4 Proficient

- displays coherent thought and organization with some evidence of categorization and/or superordination/subordination

3 Satisfactory

- displays little coherent thought and organization
- categorization barely evident

2 Limited

- categorization not evident
- poorly organized

1 Poor

- no evident organization at all
- rambling and hard to follow

0 - too little writing exists to make a judgement.

0 Insufficient

- too little writing exists for a judgement to be made

B. Applying the Criteria

Included here are twelve grade 4/5 papers on which to practice the practical application of this scoring method. Read each paper and by referring back to the descriptors assess a mark from 0-5 for each of the areas. You must remember to check the criteria for each sub-component before reading the paper to mark for that component. As a means of comparison, you can find the scores that the marking team awarded each paper by looking in the appendix section of this document.

It is important to refer always to the descriptors and mark according to how the paper meets those standards. Try not to compare papers to ones read previously.

Also, when marking for sentence structure you may have to mentally put in the correct punctuation - if the sentence structure is there the paper will be penalized in the conventions section for lack of punctuation, but shouldn´t be penalized twice.

MARKING SHEET

Paper #	General Impression	Conventions	Sentence Structure	Vocabulary	Content	Develop
327						
528						
433						
2						
541						
523						
229						
368						
207						
51						
57						
266						

I.D. NUMBER 327

(GOOD COPY OR FINAL DRAFT)

PLANNING A CAMPING TRIP

If I were to go camping for two days, I would need a very warm sleeping bag so I would be warm and comfortable during the night.

I would also bring food so I wouldn't go hungry.

And to cook the food I would bring cooking utensils, such as pots, pans, platers, plates and silverwear.

I would also bring a flashlight so I could see at night, and some extra batt

Also, in case I got lost, I would need a map and compass to get home.

Some extra cloths, would probably also be smart, incase the one's you were wearing got dirty.

If it got cold or started to rain I would need a tent.

And don't forget about the first-aid, insect reppelant and matches.

For recreational purposes I would bring fishing poles, hooks and other

I.D. NUMBER 327

(GOOD COPY OR FINAL DRAFT)

fishing things if there was a lake or stream nearby.

But if there wasn't any lake or stream nearby I would bring a note book, and binoculars to go on a nature hunt. I would bring these recreational things incase I got bored.

Oh! And instead of a gun for protection against wolves or bears, bring a watchdog along! Because camping is alot more fun with a friend!

By: [redacted]

I.D. NUMBER 528

(GOOD COPY OR FINAL DRAFT)

Camping

Hi my name is Lance. One friday after ~~school~~ noon I was walking home from school. When I got home I asked my mom what we were going to do my mom seide we are going camping O geat! then she asked if I hood help her put all the suff in the trunk of the car.

When we were finshed my dad came home, me my mom and my dad watked up we eat. after we were finshed eating we left.

When we got to spring lake we un packed our tent, fishing rods, lawn chairs, sleeping bags are gun blankte for the pick nick tabel and one flare gun after we finshed un packing and got everthing setup it was getting late so me my mon and my dad went to sleep when we woke up in the moning we aet beak-fast.

after beakfast we whent fishing for tow hours. we caght 6 fish then we ~~ate the fish~~ whent back to campe and after we ate the fish. Then we put all of the caming gear in the trunck then we slated home. And llat was the camping trip.

The End

433

I.D. NUMBER

(GOOD COPY OR FINAL DRAFT)

we paked it up. lets chek it cmpess
tent, gas stove, knif, axe, gi[illegible],
sleeping bags, Bob cat, maches, car,
gas, food, fishing poles. ok lets go.
clothing, now may we go yey yey
it will take a wold (~~dog~~) day. ron
last one in is a drie pig splas
go get the conenew l will the
fire strted

I.D. NUMBER 2

(GOOD COPY OR FINAL DRAFT)

Today is the day! We're leaving t~~oday~~ on a camping trip right after lunch! I've got everthing ready to be put in the car. My mom packed some goodies for the trip. We're all ready now, my brother as usual, is whinning about not being able to take all of his stuffed animals!

Finally we're on our way to the campsite! My dad is driving, my mom is reading, my brother is sleeping and I'm coloring. It is so peacful and quiet that soon I'm fast asleep. But two hours later I am awakened by a clap of thunder! Now it's not so peacful and quiet its pouring rain and thundering and lightening.

"Look for campsite number twenty four" my dad soon calls out. "Boy am I ever glad we're here" I said, "We're all out of goodies, all we have left are marshmallows!" "To bad its still drizzling and wet" said dad" I wanted to play some ball." "It will probably be dry by tommorow" said mom. "Hey there's number twenty four!" called out Michael." Well lets get some supper, I'm famished!" I said "Yeh me to whats for supper mom" said Michael." Hamburgers and vegetables" said mom. I giess we'll have to try and start a fire" said

I.D. NUMBER 2

(GOOD COPY OR FINAL DRAFT)

sad, "Michael you go and get some wood for a campfire, then we'll put up the tent"

"Mm Mm Are these Hamburgers ever good" I said. "Yes they're delicous" said dad "Lets go frog hunting Ali!" said Michael. "Oh alright" I said. Two hours went by and we had caught five frogs but let them go when our mom called, "Bedtime!" "We're coming" I called back.

The next morning when I got up everyone was eating breakfast "Good morning Allison" my mom said, "Good morning" I said as I sat down to breakfast. "Today is going to be a big day" dad announced "I hope you got a good sleep" Was he ever right, after breakfast our whole family played baseball, me and mom against Michael and dad. Michael and dad won 6-8, still it was alot of fun! After our game of baseball we had lunch it was dill pickles and ham sandwiches and for dessert we had wild strawberries. Then we went for a refreshing swim and a soccer game.

Then came supper it was superior! My dad cooked some steak and beans over a campfire on a grill. After supper we had a quiet evening roasting marshmallows and playing cards.

I.D. NUMBER 2

(GOOD COPY OR FINAL DRAFT)

That night we all went to bed quite late which was 11:30 so we all slept in till 9:30.

For breakfast we cooked bacon, eggs and toast on a grill over a campfire "Today we are spending a day on the lake" said dad "Hooray!" We all shouted. After breakfast mom and I packed a lunch and we towed the boat out to the lake and off we went fishing rods and all! We fished until lunch which was hot dogs and fruit. We kept fishing until 5:00 so supper would be ready by 6:00. We had a delicous supper of fish and wild strawberries.

At bedtime (which was 8:00 because it was going to be a long day tommorow) I fell asleep as soon as my head reached the pillow! Tommorow we were leaving!

I.D. NUMBER 541

(GOOD COPY OR FINAL DRAFT)

Jason and his famliy were going on a camping trip.

On the way to the camping grounds they saw a big brown bear. The big brown bear hurried to the nearest brush and hid behide it

Jason new he was going to have fun. When Jason and his dad and mom and sister got to the grounds

They helped unpack all there things. Jason and dad put up the tent.

While mom and I went to get some wild berrys to eat.

When we got back the tent was up and Jason and dad were sitting on the table ready for some wild berrys.

Mom washed them and I put them on the table after we had the berrys it was time for bed. When it was morning we packed up all are things and went home.

The End

I.D. NUMBER 523

Canping

(GOOD COPY OR FINAL DRAFT)

I am Michael Last sammer I went canping I canp at vags font my mom and Dad set up the lentitrelr well Jahn was halling yeuald I wrz coping it the neat bay I wro up bre for mam Dad and Jahn I blaiy my biny up I put the podlas in it I pad the fishing gerin ta I rot a not ta mam and Dad I Bread the Deny to the lack I fish all day I cot Jave fish. wen I gat back it was tine ta go lo bed the nest moning we had to ga.

I.D. NUMBER 229

(GOOD COPY OR FINAL DRAFT)

Teacher for a Day

In the morning I would start with the every day work we do all the time. We would have Language Arts, and read in our readers and work in our Reading workbooks. We would do that from 9:00 to 10:00. Then from 10:00 to 10:30, we would problbly correct the math work from yesterday and start with a new assiment in math. At 10:30 the children would go outside till 10:55. They would come in and I would tell them what to get out for next class, which would be Social Studies from 11:00 to 11:30. In Social Studies we would study "Should Canada Share it's Natural Resources"; a unit we are taking.

At 11:30 I would tell them to take out there reading books and read quietly, or if they had to exchange there books I would send them to the library row by row. At 11:50 the lunch bell would ring, so I would tell them to get there coats and line up at the door quietly. They would be dismissed for lunch, and I too would go home for my lunch. I would be back about five minutes before the children come back from recess. They would come in and have there next class out. I would have written it on the borad. I't would read "gym".

I.D. NUMBER 229

(GOOD COPY OR FINAL DRAFT)

I would send them to change and come back to the class room quietly and wait for me quietly in the class room. When they were all back we would go down to the gym and do stretches and play dodge ball or scocer baseball. After this they would change and come back to the class room. From 1:30 to 2:15 we would do French with a speical teacher. After French it would be recess, I would line them up and send them to the playground. The teacher "me", would mark some work they did in Language Arts or any other subject. At 2:30 it would be spelling. I would have paper on their desk to write spelling words. I would read them untill all the words are finished. At 2:45 we would have Science and do "Magnets" a nother unit we are studing. At 2:30 I would tell them to hand in the work they were doing so I could mark it.

If some kids were not done it would be homework, so that the real teacher wouldn't be mad. They would be dismissed quietly and leave. I would still stay and mark the work I didn't do. When I was finished I would put my stuff away and get my coat and shut off the lights.

I.D. NUMBER 368

(GOOD COPY OR FINAL DRAFT)

Teacher for a Day

Lori's class chose her teacher for a day. Lori was planing what she was going to do. she decided to have spelling, french, helth, recess, music, ussr, lunch, math, langage arts, recess, wrighting physical educution. There was Joel the prackal jocker that she was worried about, And what he was going to do. He gave me werd looks after school. Next Day She was up at 6:00am

I.D. NUMBER 368

(GOOD COPY OR FINAL DRAFT)

getting things reddy. Lori left at 7:30am and got there at 7:45am. Then she went down to the fodocopy machain and ran off spelling french, math, langage arts. Lori was reddy, she was still worried about Joel. It was 9:00am everybody started to come inclouding Joel with his bag full and there was know homework. Lori started them on spelling. It was french pereod then helth and recess. Then music,

I.D. NUMBER 368

(GOOD COPY OR FINAL DRAFT)

u.s.s.r and lunch."Lori said THANK GOD!" Then math, langage arts and recess. Wrighting and physical Education. It was the end of school, she said "it isn't easy being teacher for a day!"

I.D. NUMBER 207

(GOOD COPY OR FINAL DRAFT)

Techer For a Day

Foist we should do the quick quiz and then do handwriting and then we will lean all about comuters. And after we will clean our desks and get ready for home time.

I.D. NUMBER 51

(GOOD COPY OR FINAL DRAFT)

That afternoon, when she was on her way home, she was worried that Joel would do something wrong, that would make her look silly.

When she got home, she smelled the steak in the oven and forgot about Joel. She ate her supper and was going to bed.

Before she went to sleep, she remembered about Joel. She tried to think about something els, but she couldn't. Anyway, she f got to sleep. She dreamed about the next day when she would be teacher. But for some strange reson, she was sick and couldn't go! Just then she heard a ghostly call. "Lorie, wake up! You will be late for school!" She was just opening her eyes. She thought she was still dreaming, because in her dream, she was si and slepped in her dream! She jumped up. But then she remembered how tired she was and flumped back down. After a while, she glanced at her clock. It was eight thurty!

I.D. NUMBER 51

(GOOD COPY OR FINAL DRAFT)

She got dressed in a hurry. Then, had a quick bowl of cerial, brushed her teeth, and ran to school. When she got there, the last bell was ringing. She raced into school. She was in the door way now, looking around to see if Joel the pest was there. To her surprize, he wasn't there! She walked up to the teachers desk. "You may call the roll," said Mrs. Summers. Lori called out all the names in the class and only one wasn't there. "Pest." She stepped up to the frount of the class. "Now class," Lori said "I am going to hand the tests back for you to finish." Everybody groaned. Lorie handed the tests back to glum faces. When the class had started, it was Lori's job to mark the spelling tests was fun at first, but she got so board, that after a while, she almost fell asleep! She decided that taking down stories was much more to her liking. She was just getting out of her chair, when the noon bell went. She waited for the anouncements to finish, and then dismised the class.

She was on her way home, when she found that there was a huge mud puddel in frount of her. She was going to step over it when a hand reached out of the hedge and pushed her over. She fell into the puddle with a splash! Then she heard a giggle. It was Joel! "I'll get you for this!" And she did to! That afternoon, she told the principle what had happened and how Joel was jelous that he didn't get elected. So all afternoon, Joel sat in the ofice face to face with the principle. The rest of the day went fine for Lorie.

She knew that something was going on because of Joels looks after School days. But had nothing to worr about now.

The rest of the day went fine without Joel. The class pest of the year.

The End

I.D. NUMBER 57

(GOOD COPY OR FINAL DRAFT)

Thatday Its fannaly here

9:00 friday morning it's acculy happening Lori gets to teach. First the have math. "We are going to have a quiz this affternoon I hope every body studied". "Right now let's have races first person in each row up to the board; put down $\overset{14r4}{20\overline{)284}}$" right Laura

20x
84
80
4

its 14r4". Know second person in each row up put down $\overset{14r4}{10\overline{)184}}$ right Jack it's 14r4 and Lori went

10x
84
80

through the 3rd, 4th, 5th and 6th then she went and got a film she came back with math patrol Lori started the film, during the film Lori walked around finally she came to Joel's desk she found him sleeping so she stoped the film, woke Joel up sent him into the hall and turend the film back on, It was recess time when the film was done, before she let them go out side she told them to stay outsid. After recess they went on a nature walk. "Does anybody know what flower this is asked Lori"? "Yeeow"! said Joel. Joel had Just fallen into a thorn brush Lori helped him out then they went back to the school

I.D. NUMBER 57

(GOOD COPY OR FINAL DRAFT)

it was time for lunch Lori's class is selling hotdogs Lori had a drink of orange pop but she couldn't get any because somebody pluggd her straw, she knew it had to be Joel so she gave him a rubber hotdog. Know it's after lunch and there having a party. There's orange pop pepsi coke plus cake played games the Lori gave them the quiz by the time every body was done the quiz it was recess After recess they played doge ball and Joel, Betty, Annette, Nima and Jack were in the middle. First Joel was out then Jack the Bettey then anette then Nina then they went back to the classrom and put up there chairs and went home.

I.D. NUMBER 266

(GOOD COPY OR FINAL DRAFT)

One day joel put a tac on the teacher's chair. and then the teacher sent a note to Mr. Murragus. and then Mr. Murragus asked joel to come in the office and then they hada talk about his behaver. then he said you better smarten up joel, o.k o.k joel said and he said thank you for the talk Mr. Murragus, and then Troy came walking along to the Bathroom and joel went to the Bathroom to talk about what happened in the of office oh we had a talk about my behaver oh we better get back to class be for we get in trouble, from the teacher try being nicer to her o.k and think where your going to go when you die joel. oh o.k I.ll be nicer to her asa Subsl. I.ll even give her some fruit all the go sand I.ll even be nice to the kids in the school. hello said joel to Lori and Lori said hello to joel said I think I have a little pie look in your left drower. o.k joel ah hhk a frog. now look in your right hand drower. o.k better not be a frog joel no it's not a frog, joel said a box of candy thank you joel. now look in your's a ten doller Bill thank you Lori. and he gave every one something too and Mr. Murragus liked his surprise, and thanked him for it and then every body thanked him for being nice to every one in the school.

III. PRACTICAL CLASSROOM APPLICATIONS

The holistic approach to marking identified components of student writing has several practical classroom applications.

It can assist the teacher in diagnosing strengths and weaknesses both in student performance and program delivery. This scoring method provides the teacher with very useful and appropriate feedback for reporting student progress and in planning individual program strategies. Having a standard set of criteria leads to a more objective and consistent evaluation of student writing.

Teacher marking time can be more effectively utilized by providing a greater focus for the evaluation of student writing. It is not necessary to mark every writing assignment for all components. If you have, for example, taught a lesson on the use of specific versus general words, you could mark the written assignment using only the vocabulary criteria.

Another application of the criterion is for students use in evaluating their own papers. It would be possible to put the criteria in a more simplified form so that students could gain some appreciation for the elements of good writing. They could grade their own performance and see how they could improve their written composition based on the marking criteria. Although students would probably experience some problem initially, this approach could be employed to develop skills in proofreading and editing both a student´s own work or the work of others.

IV. RESULTS SUMMARY

1. (a) Overall District Results - Grade Four Written Composition

	Mean	Std. Dev.	Confidence Int.		
General Impression	2.8	.8	2.6	to	3.2
Content	3.1	.9	3.0	to	3.3
Development	3.1	.9	3.0	to	3.3
Sentence Structure	2.9	1.0	2.8	to	3.0
Vocabulary	3.0	.7	2.9	to	3.2
Conventions	2.9	.9	2.8	to	3.0
Average of 5 Components	3.0	.8	2.9	to	3.1

1. (b) Overall District Results - Grade Five Written Composition

	Mean	Std. Dev.	Confidence Int.		
General Impression	3.1	.9	3.0	to	3.2
Content	3.4	.9	3.3	to	3.5
Development	3.6	.9	3.5	to	3.7
Sentence Structure	3.3	.8	3.2	to	3.4
Vocabulary	3.4	.7	3.3	to	3.5
Conventions	3.3	.8	3.2	to	3.4
Average of 5 Components	3.4	.7	3.3	to	3.5

NOTE: By comparing class results to the above District results teachers can determine possible strengths and weaknesses in both student performance and teaching strategies.

2. Correlation Among Written Composition Variables

a) Grade 4

	Impr.	Cont.	Devl.	Sent.	Vocab.	Conv.	Averg.
Impression	1.00	.77	.76	.70	.74	.70	.85
Content		1.00	.66	.55	.76	.53	.80
Development			1.00	.79	.69	.76	.90
Sentence				1.00	.71	.83	.90
Vocabulary					1.00	.67	.87
Conventions						1.00	.88
Average of Sub-Comp							1.00

b) Grade 5

	Impr.	Cont.	Devl.	Sent.	Vocab.	Conv.	Averg.
Impression	1.00	.79	.72	.66	.74	.87	.84
Content		1.00	.66	.60	.83	.58	.86
Development			1.00	.72	.58	.73	.87
Sentence				1.00	.58	.79	.86
Vocabulary					1.00	.58	.82
Conventions						1.00	.86
Average of Sub-Comp							1.00

A correlation coefficient is an index of the linear relationship between two variables. A perfect correlation relationship is 1 (as one variable goes up the other goes up as well) but a correlation greater than .5 is considered significant.

3. (a) Percentiles for Written Composition Variables - Grade Four

Score	Impr.	Cont.	Devl.	Sent.	Vocab.	Conv.	Averg.
.7	1		1		1	1	
1.0	1	1	1	2	1	2	1
1.3	3	2	2	5	2	5	2
1.7	7	4	4	9	4	10	4
2.0	19	13	9	18	9	19	8
2.3	34	27	17	27	17	29	18
2.7	49	36	32	39	29	40	35
3.0	65	50	48	55	53	53	54
3.3	77	64	61	68	74	69	68
3.7	84	73	71	78	84	80	80
4.0	91	80	80	85	91	87	89
4.3	95	87	89	91	95	93	93
4.7	98	95	94	95	98	97	97
5.0	99	99	99	99	99	99	99

3. (b) Percentiles for Written Composition Variables - Grade Five

Score	Impr.	Cont.	Devl.	Sent.	Vocab.	Conv.	Averg.
1.3	2	1	1	1		1	
1.7	5	3	3	3		3	1
2.0	12	8	5	5	2	7	3
2.3	21	16	9	11	6	13	7
2.7	33	23	17	21	14	24	17
3.0	50	35	27	34	32	38	32
3.3	64	49	37	50	52	55	48
3.7	74	61	49	67	68	68	65
4.0	83	70	64	80	80	78	79
4.3	88	79	78	88	90	88	88
4.7	94	89	89	93	95	95	95
5.0	99	99	99	99	99	99	99

4. Reliability Study - Written Composition Marking

Grade	Scale	All Agrees	All Within 1	Disagree
Three	General Imp.	32%	90%	10%
	Content	30%	88%	12%
	Development	21%	80%	20%
	Sentence	23%	83%	17%
	Vocabulary	32%	91%	9%
	Conventions	34%	96%	4%
Four	General Imp.	32%	87%	13%
	Content	34%	89%	11%
and	Development	24%	78%	22%
	Sentence	25%	80%	20%
Five	Vocabulary	39%	90%	10%
	Conventions	24%	81%	19%

Dr. Tom Maguire of the University of Alberta carried out the above inter-rater reliability study. Essentially, this is a statistical measure of the amount of agreement or disagreement among the raters who scored the papers. This type of study sheds light on the question of whether, in this type of scoring, there is adequate agreement among raters that the process can be considered reliable. Dr. Maguire's conclusion was that in terms of reliability, the judges or raters had done an excellent job. He suggested that it would be safe to use two raters, and in the event of disagreement, a third head scorer could cast a deciding vote.

5. Analysis of Variance Grade 4/5 Written Composition

(Scores are listed out of 15 possible)

Variable	Source	df	Mean Square	E	Sig	
Gen. Imp	Grade (A)	1	132.04	19.97	yes	Gr.4: 8.3
	Topic (B)	1	18.71	2.83		Gr.5: 9.2
	Task (C)	1	50.48	7.64	yes	Expository: 8.4
	AB	1	11.44	1.73		Narrative : 9.0
	AC	1	2.01	.30		
	BC	1	44.94	6.80	yes (see A below)	
	ABC	1	.54	.08		
	Within	547	6.61			

Content	Grade (A)	1	131.54	16.83	yes	Gr.4: 9.2
	Topic (B)	1	2.31	.30		Gr.5: 10.2
	Task (C)	1	6.02	.77		
	AB	1	2.24	.29		
	AC	1	.97	.12		
	BC	1	147.40	18.86	yes (see B below)	
	ABC	1	1.89	.24		
	Within	547	7.81			
Develop.	Grade (A)	1	239.44	33.87	yes	Gr.4: 9.4
	Topic (B)	1	79.06	11.19	yes	Gr.5: 10.8
	Task (C)	1	.09	.01		Teacher: 10.5
	AB	1	11.58	1.64		Camping: 9.8
	AC	1	19.03	2.69		
	BC	1	13.18	1.86		
	ABC	1	10.31	1.46		
	Within	547	7.07			
Sentence	Grade (A)	1	191.69	29.02	yes	Gr.4: 8.8
	Topic (B)	1	213.08	32.26	yes	Gr.5: 10.0
	Task (C)	1	274.96	41.63	yes	Teacher: 10.1
	AB	1	16.15	2.45		Camping: 8.8
	AC	1	4.28	.65		Expository: 8.6
	BC	1	13.42	2.03		Narrative :10.0
	ABC	1	8.92	1.35		
	Within	547	6.61			
Vocab.	Grade (A)	1	197.72	43.82	yes	Gr.4: 8.9
	Topic (B)	1	27.76	6.38	yes	Gr.5: 10.1
	Task (C)	1	35.59	8.18	yes	Teacher: 9.8
	AB	1	7.32	1.69		Camping: 9.3
	AC	1	.36	.08		
	BC	1	58.77	13.50	yes (see C below)	
	ABC	1	.77	.16		Expository: 9.2
	Within	547	4.35			Narrative : 9.7
Conven.	Grade (A)	1	173.65	25.04	yes	Gr.4: 8.7
	Topic (B)	1	99.92	14.41	yes	Gr.5: 9.8
	Task (C)	1	17.48	2.52		Teacher: 9.7
	AB	1	24.52	3.54		Camping: 8.8
	AC	1	9.34	1.35		
	BC	1	9.29	1.34		
	ABC	1	.88	.13		
	Within	547	6.93			

Aver. of 5	Grade (A)	1	183.66	37.60	yes	Gr.4: 9.0
	Topic (B)	1	55.56	11.37	yes	Gr.5: 10.2
	Task (C)	1	347.54	7.11	yes	Teacher: 9.9
	AB	1	10.41	2.13		Camping: 9.3
	AC	1	2.58	.53		
	BC	1	35.40	7.25	yes (see D below)	
	ABC	1	.44	.09		Expository: 9.3
	Within	547	4.89			Narrative : 9.8

In all cases, the averages for grade 5 were significantly higher than grade 4. In those cases where "Topic" was significant, the "Teaching" mean was higher than the "Camping" mean. In those cases where "Task" was significant, the "Narrative" mean was higher than the Expository mean.

The BC interaction was significant in the cases of: General Impression, COntent, Vocabulary, and Average of 5 Components. Generally, this resulted from an unexpectedly high value for the narrative papers written on the topic of teaching.

	A Impression		B Content		C Vocabulary		D 5 Sum.	
	Exp.	Nar.	Exp.	Nar.	Exp.	Nar.	Exp.	Nar.
Teach.	8.2	9.5	8.7	10.2	9.0	10.3	9.3	10.4
Camp.	8.5	8.6	10.0	9.4	9.3	9.3	9.2	9.3

V. APPENDIX A - ANSWER GUIDE

Following are the marks that the marking team gave each paper.

MARKING SHEET

Paper #	General Impression	Conventions	Sentence Structure	Vocabulary	Content	Develop
327	4	5	4	4	4	5
528	3	3	3	3	2	4
433	1	1	2	3	2	2
2	5	5	5	5	5	5
541	2	4	4	3	2	4
523	1	0	1	1	1	2
229	5	4	4	4	4	5
368	3	3	3	4	2	3
207	2	1	1	2	1	3
51	4	4	5	5	5	5
57	3	2	3	3	3	3
266	2	2	3	3	2	2

V. APPENDIX B

GENERAL IMPRESSION SCORING DESCRIPTORS FOR DESCRIPTIVE WRITING MODE

5 Excellent

- exceptionally developed detail
- precise vocabulary
  - vocabulary displays a variety of new and interesting words
  - comparisons create vivid impressions
- displays exceptional thought and organization
- shows some evidence of style
- a dominant impression may be evident
- few mechanical errors
  - punctuation
  - sentence structure
  - capitalization
  - spelling

4 Very Good

- well developed detail
- some precise vocabulary
- displays good evidence of thought and organization
- some mechanical errors - but these don't interfere with readability or meaning
- also includes Number 5 papers with many mechanical errors.

3 Average

- displays some thought and organization
- sufficient detail
- length is adequate to complete the task
- uses vocabulary appropriate to grade level
- mechanical errors interfere somewhat with the message and the readability.

2 Poor

- length is inadequate to complete the task
- lack of detail
- vague vocabulary - dull uninteresting words
- lack of thought and organization
- overall impression of disorder due to jumbled arrangement of ideas
- irrelevant details
- mechanical errors greatly interfere with message

1 Little or No Communication

- mechanical errors interfere with meaning to the extent that the composition is nearly illegible
- completely off topic

0 Insufficient

- too little writing exists for judgement to be made

CONTENT SCORING DESCRIPTORS
FOR DESCRIPTIVE WRITING MODE

5 Exceptional

- specific details used to describe setting and activities
- creates an atmosphere through the use of the senses
- creates a vivid overall impression and gives clear physical descriptions
- good use of imagery
- captures a dominant impression or sense of style

4 Proficient

- some specific details to describe setting
- some general sense of atmosphere is created
- appeals to most of the senses
- use of imagery is evident

3 Satisfactory

- evidence of specific appropriate details
- some attempt to create an atmosphere or overall impression
- attempts to use imagery

2 Limited

- few appropriate details
- very little attempt to create atmosphere
- limited appeal to senses
- very little use of imagery

1 Poor

- no appropriate details
- setting is not developed
- no sense of atmosphere
- no imagery

0 Insufficient

- too little writing exists for judgement to be made

DEVELOPMENT SCORING DESCRIPTORS FOR DESCRIPTIVE WRITING MODE

5 Exceptional

- displays coherent thought and organization
- there is evidence of paragraphing
- organized sequence of descriptions and events
- shows excellent sense of beginning and closure

4 Proficient

- displays good evidence of thought and organization
- good sense of beginning and closure
- may have some slight confusion in flow of ideas

3 Satisfactory

- descriptions are in generally coherent sequence
- some sense of closure is evident
- some disorganization of ideas

2 Limited

- limited sense of sequencing the descriptions
- absence of sense of closure
- weak sense of organization

1 Poor

- no sequencing
- no closure
- no evidence of organization

0 Insufficient

- too little writing exists for judgement to be made

SENTENCE STRUCTURES SCORING DESCRIPTORS
FOR DESCRIPTIVE WRITING MODE

5 Exceptional

- good variety of sentence structures, type, length is used
- controlled used of co-ordination
- sentence fragments if evident are used for effect

4 Proficient

- some variety in sentence structure, type and length
- little over-use of co-ordination
- few sentence fragments

3 Satisfactory

- little variety in sentence structure, type and length
- some over-use of co-ordination
- some sentence fragments evident

2 Limited

- most sentences are simple sentences
- little variety in length and structure
- definite use of co-ordination
- may have many sentence fragments

1 Poor

- sentences are immature and repetitious
- almost exclusive use of co-ordination
- sentence fragments impede meaning

0 Insufficient

- too little writing exists for a judgement to be made

VOCABULARY SCORING DESCRIPTORS
FOR DESCRIPTIVE WRITING MODE

5 Exceptional

- specific, concrete, interesting words have been selected to create vivid images and precise details
- denotative meanings are accurate and effective

4 Proficient

- frequent use of specific concrete words adds clarity to the detail created
- denotative meanings are most frequently accurate and effective

3 Satisfactory

- some use of specific, concrete words
- some use of general words
- denotations are mostly correct

2 Limited

- few specific concrete words
- most words are general
- some inaccuracy of meaning

1 Poor

- only vague, general words are used
- restricted choice of words
- inaccuracy of meaning

0 Insufficient

- too little writing exists for judgement to be made

CONVENTIONS SCORING DESCRIPTORS
FOR DESCRIPTIVE WRITING MODE

5 Exceptional

- the communicative power of the composition is enhanced because of careful form, spelling, usage, punctuation and capitalization and neatness of writing.

4 Proficient

- communication is clear because of essentially correct form, spelling, usage, punctuation and capitalization
- few errors in proportion to length

3 Satisfactory

- some errors in form, spelling, punctuation, usage and capitalization but communication is adequate

2 Limited

- frequent errors in spelling, punctuation, usage and capitalization reduce communication
- work is not neatly done

1 Poor

- very weak in communication due to incorrect spelling, no punctuation and capitalization and poor grammar
- poor printing/writing make it very hard to read

0 Insufficient

- too little writing exists for a judgement to be made

V. APPENDIX C - STUDENT BLOOPERS

- In health, the bodies were coming along real well.
- Beans we are going to bring for fast travel.
- Susan thought for awhile, but she couldn´t because the photocopier was broken.
- And this one word I´m about to write will scare the hell out the bugs around the camp - RAID!
- She told her class to get ready for P.G.
- Walking through the flowers at "Bushard" Gardens.
- He asked Lori how the kids were
  "Like angels" she replied ...
  "Well, maybe Hell´s Angels"
- If I give the kids recess they will respect me more.
- At 12:30 we get noon.
- They spread the tent out and put the pigs in the ground.
- Then it was summer holidays and Lori moved to Singapore and got eaten by a snake.
- Lori Picklehoper

GRANDE PRAIRIE SCHOOL DISTRICT #2357

LANGUAGE ARTS PRODUCT EVALUATION

WRITTEN COMPOSITION ASSESSMENT

GRADE FOUR/FIVE

June 1984

TEACHER'S MANUAL

ADMINISTRATION DIRECTIONS

I. Background

This test of written composition is part of the Grande Prairie School District Language Arts Product Assessment Research Project. One of the research questions being investigated is whether students write as well in the expository persuasive mode as they do in the narrative mode. Consequently, students in your class have been randomly assigned to do either Writing Task A (expository/persuasive) or Writing Task B (narrative). Besides answering the above research question, other side benefits of this project include the following:

A. You will get some feedback regarding how well students in your class performed in the various skills that make up written composition, compared to the performance of a larger sample of students. This information may be valuable to you in making future decisions regarding your instructional emphasis.

B. Those teachers who participate in the scoring teams will gain experience in holistic methods of scoring. In addition, they will produce a holistic scoring handbook which might prove useful to you.

This test is meant to be administered over a two day period. This is to provide as realistic and desirable writing situation as possible. Students do their best writing when they have been provided some time to think about their ideas, and to discuss them with their peers.

II. Day One Instructions

In a double period block (approximately 70 to 90 minutes) hand out the test booklets and read through the two writing topics with the students. Explain that they may choose either topic, but regardless of the topic, they will be assigned by you to Writing Task A or Task B and will have no choice in that matter. Give them about 5 minutes to reread the topics, at the end of which they will be asked to select a topic. Then divide the class into groups of approximately 3 to 6 students who have chosen the same topic and have been assigned to the same Writing Task. Tell the students that they will be given 15 minutes to discuss and share ideas about their Topic-Task. They are not to do any writing at this time. At the end of the 15 minutes, reorganize the class so that students work individually for the remainder of the period. (Approximately 40 minutes). Tell them that they can begin writing their compositions now, and will be given time to complete them and/or write a second draft the following day.

At the end of the period, collect their work to that point. They are not to be allowed to take their work home with them.

III. Day Two Instructions

On Day Two, students should be provided two consecutive class periods (from 70 to 80 minutes) in which to either complete their first drafts, edit, and revise, or write a second draft. Provide an alternative quiet activity for those who finish early.

The following guidelines should be shared with the students:

1. There is no required length, but marking consideration will be given to the extensiveness or completeness of ideas expressed.

   The composition can be one or more paragraphs, depending on what they feel is necessary to express their ideas in an organized way.

2. Compositions should be handwritten, not printed.

3. Pens should be used - not pencils.

4. Students cannot use dictionaries or other aids, and the teacher (or friends) is not allowed to supply words, spellings, or other assistance.

IV. Shipping Instructions

1. Make sure that students have filled in the required information on the front page of the test booklet, and that the booklets are stapled together securely with the pages in order. Students should not write their names anywhere on their compositions - only on the front page of the test booklet.

2. Bundle the test booklets up by class and forward them to Keith Wagner before June 30.

Thank you for your cooperation.

V. Additional Information

My memo of February 6 and information regarding scoring techniques are appended.

PREAMBLE - In July 1984, nine teachers worked as a team to evaluate approximately 550 Grade 4/5 written compositions. Half of the compositions were narrative; half were expository/persuasive. The first task was to assess each composition for overall general impression on a scale of 0 to 5. The team also developed descriptors for each of 5 sub-components: content, development, sentence structure, vocabulary, and conventions. All of the descriptors are outlined below:

GENERAL IMPRESSION SCORING DESCRIPTORS

5 Excellent

- exceptional clarity of communication
- has an evident, developed style
- creativity and specificity of detail which is suited to the purpose and relevant to the topic, including the stating of most reasons in the persuasive composition.
- vocabulary is specific, descriptive, vivid, and connotative
- exceptional thought and organization, including an evident beginning, middle, and ending
- very few errors of convention relative to the length

4 Very Good

- clarity evident, but atmosphere or style may not be found consistently throughout the paper
- appropriate amount of detail, including the use of many reasons in the persuasive composition some precise vocabulary
- displays good evidence of thought, and a beginning, middle, ending sequence
- some mechanical errors, but not so many that they interfere with readability or meaning

(this category may also include #5 excellent compositions that have many mechanical errors)

3 Average

- communicates satisfactorily, length is adequate to complete task
- satisfactory detail suited to purpose, including an attempt made at giving details in a persuasive paragraph
- uses vocabulary appropriate to topic and purpose
- mechanical errors interfere somewhat with the message and the readability.

2 Weak

- length is inadequate to complete the task
- lack of sufficient detail, inclusion of some irrelevant details
- creates an overall impression of disorder, and lacks a clear ending
- vague vocabulary
- mechanical errors greatly interfere with meaning

1 Poor

- length is inadequate to complete task
- cannot tell what the purpose or task is
- very few details and several of them are irrelevant
- mechanical errors interfere with meaning to the extent that the composition is nearly illegible

0 - No real communication at all

CONVENTIONS - (Narrative and Persuasive Task)

5 Exceptional

- communicative power is enhanced because of careful spelling, grammar, punctuation, and capitalization

4 Proficient

- communication is clear because of essentially correct spelling, grammar, punctuation, and capitalization

3 Satisfactory

- communication is adequate because of generally correct spelling, grammar, punctuation, and capitalization

2 Limited

- communicative power is reduced because of incorrect spelling, grammar, punctuation, and capitalization

1 Poor

- communicative power is very weak because of errors in spelling, grammar, punctuation, and capitalization

SENTENCE STRUCTURE - (Narrative and Persuasive Task)

5 Exceptional

- variety of sentence type, length, and structure is used for effects such as emphasis
- coordination has been controlled, and subordination is used appropriately
- sentence fragments, if present, are used for effect

4 Proficient

- variety is evident
- coordination is seldom overused
- subordination is often used appropriately
- there are few inadvertent fragments

3 Satisfactory

- some variety evident, but coordination may be overused
- subordination is successfully attempted
- fragments are in evidence, but do not impede meaning

2 Limited

- little variety and some awkward structures
- overdependence on coordination
- subordination, if used, is inappropriate
- fragments are frequent and impede meaning

1 Poor

- sentences are immature and there are many repetitious patterns
- coordination is used almost exclusively
- fragments are common and impede meaning

VOCABULARY - (Narrative and Persuasive Task)

5 Exceptional

- occasional specific concrete words selected to create vivid images or precise details
- word meanings are accurate and effective

4 Proficient

- some use of specific or concrete words adds clarity to the detail created
- most word meanings are accurate and effective

3 Satisfactory

- some words have been selected appropriately but general or abstract words are often used where specific or concrete words would have been more effective
- some word meanings may be inaccurate or ineffective

2 Limited

- general words are usually used where some specific words would have been more effective
- many word meanings may be inaccurate or ineffective

1 Poor

- words convey only vague or general meanings

CONTENT - (Narrative Task)

5 Exceptional

- events are plausible and consistent with purpose
- specific details develop character, setting, atmosphere, or events

4 Proficient

- events are plausible
- appropriate details present a description of characters, setting, or events

3 Satisfactory

- most events are plausible
- several details used to describe characters, setting, or events

2 Limited

- many events are plausible
- a few details are used to describe characters, setting, or events

1 Poor

- events are implausible
- no details used to describe characters, setting, or events

CONTENT - (Persuasive Task)

5 Exceptional

- choices and reasons are plausible
- paper represents exceptional thought

4 Proficient

- choices are plausible, and reasons are given for most choices
- paper represents a good deal of thought

3 Satisfactory

- most choices are plausible, and some are supported with reasons
- paper reflects some thought

2 Limited

- some choices are plausible
- very few or no reasons given for choices
- paper represents little thought

1 Poor

- most choices are implausible
- no reasons given to support choices
- no real thought evident in the paper

DEVELOPMENT - (Narrative Task)

5 Exceptional

- displays coherent thought and organization
- supported to some degree by paragraphing and/or by transitionals (before, after this, meanwhile)
- contains organized sequence of description and events, including excellence of beginning, middle, and closure

4 Proficient

- displays coherent thought and organization
- shows distinct beginning, middle, and end

3 Satisfactory

- generally coherent and organized
- both beginning and closure are evident

2 Limited

- some lack of coherence and organization
- closure may not be evident

1 Poor

- no evident organization at all
- rambling and hard to follow

DEVELOPMENT - (Persuasive Task)

5 Exceptional

- displays coherent thought and organization, with obvious evidence of categorization and/or superordination/subordination
- supported to some degree by paragraphing and/or by transitionals (for example, because of this, etc.)
- includes evidence of introduction and closure

4 Proficient

- displays coherent thought and organization with some evidence of categorization and/or superordination/subordination

3 Satisfactory

- displays little coherent thought and organization
- categorization barely evident

2 Limited

- categorization not evident
- poorly organized

1 Poor

- no evident organization at all
- rambling and hard to follow

I.D. NUMBER ☐

GRANDE PRAIRIE SCHOOL DISTRICT #2357

WRITTEN COMPOSITION TEST

GRADE FOUR/FIVE

June 1984

STUDENT BOOKLET

STUDENT'S NAME ______________________________

SCHOOL ______________________________

TEACHER'S NAME ______________________________

WRITING TOPIC - Circle One
Camping
Teacher For A Day

WRITING TASK - Circle One
Task A
Task B

INSTRUCTIONS TO STUDENTS

1. This is a test to see how well you can express your ideas in writing. Do your very best.

2. There are two topics. You will be given your choice of one of the two topics.

   Each topic has two possible writing tasks. Your teacher will assign you to either Task A or Task B.

3. You will be given a chance to discuss your topic with other students for a few minutes before you start writing. You will start your writing, let it set over night, and complete a second draft the next day.

4. Your work should be done:

   (1) In handwriting, not printing.
   (2) In pen, not pencil.

5. Once you start writing, you won´t be allowed to use a dictionary or other books, or to get help from your teacher or other students.

TOPIC I: CAMPING

Jason smiled broadly. It was the last hour of the last afternoon of the last day of school. Tomorrow he would be going camping with his mom and dad and sister. Jason thought of the fun they would have on this camping trip - a real camping trip with a tent and sleeping bags. They would even be cooking outside on a real fire.

Jason and his father were going to take their fishing poles along. Jason's mother and sister had been talking all week about the wild strawberries that grew in the open places in the woods near their camping spot. Jason's mouth watered as he thought of fresh trout cooked over an open fire and wild strawberries for dessert.

And there would be plenty for them all to do. They would be able to swim in the lake or ride horses from the stable down the road. Even walking through the woods watching for animals would be fun.

Jason thought of all of the possible adventures he could have exploring the woods with his sister. His smile increased. He could hardly wait for tomorrow to come.

Your teacher will assign you one of the tasks below. For your assigned task, write a composition that is well-organized and contains a variety of words, phrases, and sentences. Space is provided in this booklet for a first draft and a final copy.

TASK A

If you were planning a camping trip, what types of things would you want to take with you? Remember that you would need shelter, food, and clothing on you trip. Also, you might wish to include gadgets that would be useful for special purposes. Imagine that you would be camping for two days and could take only what you could put in the trunk of a car. Give reasons why you would take the things that you include.

TASK B

Write a story that tells what actually happens on Jason's camping trip, or on a camping trip that you went on. The story does not have to be true. Remember that an interesting story has a beginning, middle, and end. Also, a good story tells where the action took place, and who the story is about.

TOPIC II: TEACHER FOR THE DAY

All of the members of Mrs. Summer's Grade 6 class were looking forward to the following day, but no one was quite as excited as Lori. Tomorrow was the day that Lori was to be "teacher for a day". Her classmates had elected her for this prestigious position, and Lori had spent most of the afternoon planning tomorrow's activities with Mrs. Summer while the other children worked on their language arts assignment. Everything was ready now, and it looked like tomorrow would be a good day. The only thing that worried Lori was that Joel, the class trouble-maker and practical joker, had been giving her funny looks after class.

Your teacher will assign you one of the tasks below. For your assigned task, write a composition that is well-organized and contains a variety of words, phrases, and sentences. Space is provided in this booklet for a first draft and a final copy.

TASK A

If you were elected "teacher for a day", what kinds of activities would you plan for your classmates? Give good reasons for including the activities that you have chosen.

TASK B

Write a story that tells what happens on the day that Lori is "teacher for a day". Remember that a good story has a beginning, middle, and end, and tells who the story is about, and where the story takes place.

I.D. NUMBER ☐

(GOOD COPY OR FINAL DRAFT)

I.D. NUMBER ☐

(ROUGH COPY OR FIRST DRAFT)